



Essays in Law and Economics

Citation

Yang, Crystal Siming. 2013. Essays in Law and Economics. Doctoral dissertation, Harvard University.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:11158261>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Essays in Law and Economics

A dissertation presented by

Crystal Siming Yang

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Economics

Harvard University Cambridge, Massachusetts

May 2013

©2013 – Crystal Siming Yang

All rights reserved.

Essays in Law and Economics

ABSTRACT

This dissertation consists of three essays in the area of Law and Economics.

The first essay explores the effect of judicial discretion on racial disparities in federal sentencing after the Federal Sentencing Guidelines were struck down in *United States v. Booker* (2005). Using data on the universe of federal defendants, I find that black defendants are sentenced to almost two months more in prison compared to their white counterparts, a 4% increase. To identify the sources of racial disparities, I construct a dataset linking judges to over 400,000 defendants. Exploiting the random assignment of cases to judges, I find that racial disparities are greater among judges appointed after *Booker*. Prosecutors also respond to increased judicial discretion by charging black defendants with longer mandatory minimums.

The second essay estimates the impact of increased judicial discretion on inter-judge disparities, another potential source of unwarranted disparity in the federal criminal justice system. Relying on the random assignment of cases to judges, I find that inter-judge disparities have increased significantly after the Guidelines became advisory using a random effects model. A defendant who is randomly assigned a one standard deviation “harsher” judge in the district court received a 2.6 month longer prison sentence before *Booker*, but received a 5.3 month longer sentence following *Booker*.

The third essay analyzes the differential impact of state versus federal regulation of OSHA programs on the use of traditional enforcement tools, nonfatal injury rates and fatalities, as well as wages and employment. I find that certification of a state regulated OSHA program leads to an increased use of inspections per capita and issuance of violations per capita. State regulated programs that adopted stringent standards beyond the federal program are associated with significantly lower fatalities, compared to all other states. A case study on the 2000 California penalty reform reveals that greater magnitude of sanctions and prosecution are potentially effective enforcement tools in promoting greater workplace safety and reducing fatalities. Finally, I find that there is a compensating differential for workplace safety.

TABLE OF CONTENTS

<i>Abstract</i>	iii
<i>Acknowledgements</i>	vi
<i>1. Introduction</i>	1
<i>2. Free At Last? Judicial Discretion and Racial Disparities in Federal Sentencing</i>	3
2.1. Introduction	3
2.2. Legal Background	5
2.2.1. Adoption of the United States Sentencing Guidelines	5
2.2.2. Challenges to the Mandatory Guidelines Regime	8
2.3. Data	9
2.3.1. United States Sentencing Commission	9
2.3.2. Transactional Records Access Clearinghouse	11
2.3.3. Federal Judicial Center	12
2.4. Conceptual Framework of Judicial Sentencing	12
2.5. Empirical Methodology	14
2.5.1. Potential Threats to Identification	15
2.6. Results	18
2.6.1. Sentence Length	18
2.6.2. Departures from the Guidelines	24
2.6.3. Robustness Checks for Increasing Racial Disparities	26
2.6.4. How Constraining is Appellate Review? Evidence from <i>Rita, Gall, and Kimbrough</i>	28
2.6.5. Free at Last? Effects of Judicial Sentencing Philosophies and Experience	32
2.6.6. Response of Prosecutors to Increased Judicial Discretion	44
2.7. Conclusion	45
<i>3. Have Inter-Judge Sentencing Disparities Increased in an Advisory Guidelines Regime? Evidence From Booker</i>	48
3.1. Introduction	48
3.2. Brief Legal Background of Federal Sentencing	53
3.2.1. Adoption of the Federal Sentencing Guidelines	53
3.2.2. Challenges to the Mandatory Guidelines Regime	57
3.3. Framework, Data, and Methods	59
3.3.1. Judicial Behavior in Criminal Sentencing	59
3.3.2. Sentencing Data	61
3.3.3. Matching	63
3.3.4. Testing for Random Assignment	63
3.3.5. Trends in Sentencing	66
3.3.6. Measuring Inter-Judge Disparity - Analysis of Variance	69
3.4. Results on Inter-judge and Regional Disparities	71
3.4.1. Sentence Length	71
3.4.2. Below Range Departures	74
3.4.3. Above Range Departures	77
3.4.4. Sentencing Practices by Judge Demographics	79
3.4.5. Regional Disparity: Inter-District Variation	84
3.4.6. Prosecutorial Contributions to Disparities	88
3.5. Policy Recommendations	90
3.5.1. “Topless” Guidelines System	90
3.5.2. “ <i>Blakely</i> -ized” Guidelines	91
3.5.3. Sessions Proposal	92
3.6. Conclusion	95
<i>4. The Impact of Federal and State OSHA Programs on Workplace Safety, Wages and Employment</i>	96

4.1. Introduction	96
4.2. Literature Review	99
4.3. Theory	100
4.3.1. Regulation	101
4.3.2. Enforcement	101
4.3.3. Federal versus State Enforcement	101
4.3.4. Differences in Penalty Structures	105
4.3.5. Incentive Programs	106
4.4. Data	106
4.5. Empirical Methodology	108
4.5.1. Impact of State OSHA programs on Inspections and Violations	108
4.5.2. Impact of State OSHA programs on Nonfatal and Fatal Injuries	109
4.5.3. Impact of Changes in Penalty Structure	110
4.5.4. Impact of State OSHA Programs on Wages and Employment	110
4.6. Results - Impact of State OSHA programs on Inspections Per Capita and Violations Per Capita	110
4.6.1. Differential Impact of State OSHA Programs	119
4.7. Results - Impact of State OSHA programs on Injuries	125
4.7.1. Differential Impact of State OSHA Programs	126
4.8. 2000 California Penalty Change	130
4.9. Results - Impact of State OSHA programs on Wages and Employment	136
4.10. Conclusion	142
<i>Bibliography</i>	143

ACKNOWLEDGMENTS

I am deeply grateful for the exceptional guidance provided by my thesis advisors: Lawrence Katz, Claudia Goldin, and Steve Shavell. Each has provided examples of integrity, curiosity, and dedication that I will cherish for the rest of my career. I hope this and future work is worthy of their inspiration.

I thank Gary Chamberlain, Raj Chetty, Will Dobbie, Judge Nancy Gertner, Christopher Jencks, Louis Kaplow, Ilyana Kuziemko, Jeff Miron, Alison Morantz, Shayak Sarkar, and Kate Stith for many helpful comments and suggestions, and participants in the Harvard University Labor Economics/Public Finance Seminar and Law and Economics Seminar. The Transactional Records Access Clearinghouse (TRAC) at Syracuse University, in particular Sue Long, generously provided sentencing data for use in this project in my role as a TRAC Fellow of the Center. I also thank the Olin Center for Law and Economics, and the Harvard University Program in Inequality and Social Policy for financial support.

Thank you Will for inspiring me and keeping me sane.

And finally, thank you Mom and Dad for everything. This dissertation is dedicated to you.

1. INTRODUCTION

This dissertation consists of three papers relating to the field of Law and Economics. The first two papers examine the impact of increased judicial discretion on both racial disparities and inter-judge disparities in the federal criminal justice system. The third paper analyzes the effects of OSHA programs on workplace safety, wages, and employment. The common thread throughout this work is a focus on how legal actors and institutions affect substantive outcomes of individuals.

The first paper explores the effect of judicial discretion on racial disparities in federal sentencing. The Federal Sentencing Guidelines were created to reduce unwarranted sentencing disparities among similar defendants. In this paper, I analyze the impact of increased judicial discretion on racial disparities in sentencing after the Guidelines were struck down in *United States v. Booker* (2005). Using data on the universe of federal defendants, I find that black defendants are sentenced to almost two months more in prison compared to their white counterparts, a 4% increase. To identify the sources of racial disparities, I construct a dataset linking judges to over 400,000 defendants. Exploiting the random assignment of cases to judges, I find that racial disparities are greater among judges appointed after *Booker*, suggesting acculturation to the Guidelines by more experienced judges. Prosecutors also respond to increased judicial discretion by charging black defendants with longer mandatory minimums.

The second paper estimates the impact of increased judicial discretion on inter-judge disparities, another potential source of unwarranted disparity in the federal criminal justice system. Using the judge-defendant linked dataset from the first paper, I rely on the random assignment of cases to judges. I find that inter-judge disparities have increased significantly after the Guidelines became advisory using a random effects model. A defendant who is randomly assigned a one standard deviation “harsher” judge in the district court received a 2.6 month longer prison sentence before *Booker*, but received a 5.3 month longer sentence following *Booker*. Inter-judge disparities exist in above range and below range departures. Some of the recent increase in disparities can be attributed to differential sentencing behavior by judge demographics. The application of mandatory minimums by prosecutors is another prominent source of disparities, potentially through the use of superseding indictments.

The third paper analyzes the differential impact of state versus federal regulation of OSHA programs on the use of traditional enforcement tools, nonfatal injury rates and fatalities, as well as wages and employment. I find that certification of a state regulated OSHA program leads to an increased use of inspections per capita and issuance of violations per capita. However, in the more recent period from 1996-2008, state regulated OSHA programs have no significantly lower rate of total nonfatal injuries, compared to federally regulated programs. Disaggregating state regulated OSHA programs, I find that state regulated programs that adopted stringent standards beyond the federal program are associated with significantly lower fatalities, compared to all other states. A case study on the 2000 California penalty reform reveals that greater magnitude of sanctions and prosecution are potentially effective enforcement tools in promoting

greater workplace safety and reducing fatalities. Finally, I find that there is a compensating differential for workplace safety, as wages fall significantly following certification of a state regulated program, with no changes in employment.

2. FREE AT LAST? JUDICIAL DISCRETION AND RACIAL DISPARITIES IN FEDERAL SENTENCING

2.1. Introduction

Sentencing disparities by race, gender, education, and socioeconomic status are prevalent in the federal criminal justice system. Black defendants are sentenced to five months longer in prison than white defendants who commit similar offenses and have similar observable demographic traits and criminal history. Male defendants are sentenced to over five months more in prison than similar female defendants, and defendants with lower educational attainment and income receive significantly longer sentences than otherwise similar offenders (Mustard 2001). Even within the same court, judges appear to vary significantly in their treatment of defendant race (Abrams et al. 2012), suggesting that racial disparities in the the criminal justice system may be a source of the overrepresentation of blacks in the prison population.

In response to concerns that judges were introducing unwarranted disparities in sentencing (Frankel 1973), Congress adopted the United States Sentencing Guidelines (Guidelines) under the Sentencing Reform Act (SRA) of 1984. While the Guidelines reduced inter-judge sentencing disparities in its early years (Anderson, Stith, and Kling 1999), it was criticized for its rigidity (e.g., Freed 1992 and Stith 2008), and for shifting power to prosecutors in their charge and plea bargaining decisions (Stith and Cabranes 1998, Alschuler 1978, Nagel and Schulhofer 1992).

After almost two decades of mandatory Guidelines sentencing, the Guidelines were struck down in *United States v. Booker*, 543 U.S. 220 (2005). *Booker* greatly increased the degree of judicial discretion afforded to judges (See, e.g., Berman 2005), with subsequent cases further increasing judicial discretion by reducing the degree of appellate scrutiny. Empirical work on the impact of *Booker* suggests increases in inter-judge sentencing disparities (Scott 2010), but has yielded mixed results on racial disparities, with some researchers finding large racial disparities in the aftermath of *Booker* (United States Sentencing Commission 2006, 2010) and others finding no significant impact on racial disparities in sentence length (Ulmer et al. 2010).¹ Some scholars have even argued that judicial discretion may actually mitigate recent increases in racial disparities (Fischman and Schanzenbach, forthcoming). In light of possible evidence of increasing disparities post *Booker*, the United States Sentencing Commission and policymakers have considered possible ways to constrain judicial discretion, such as “resurrecting” the mandatory Guidelines.

This paper estimates the impact of increased judicial discretion via *Booker* on racial disparities in federal sentencing using data on the universe of defendants sentenced between 1994-2009. I use a differences-in-differences (DD) methodology to compare the sentence disparities between similar defendants within a district court before and after *Booker* and find that racial disparities increase significantly after *Booker* controlling for extensive offender and crime characteristics. The black-white sentencing gap increases by almost 2 months in the post *Booker* period, a 75%

¹Both studies fail to account for district court differences, interactions between offender criminal history and offense severity, and condition on endogenous decisions to deviate from the Guidelines, which explain a large portion of increased racial disparities after *Booker*.

increase in the baseline racial gap, and a 4% increase in the average sentence length. Increased racial disparities in sentence length can be attributed to black defendants being more likely to be sentenced above the Guidelines recommended range, and less likely to be sentenced below the Guidelines recommended range, compared to similar white offenders. Even conditional on being sentenced within the Guidelines range, black defendants receive significantly longer sentences than similar white defendants. The results are robust to controlling for different racial trends in sentencing outcomes, and changes induced by other laws and court decisions. Racial disparities in sentencing persist after accounting for differential treatment of offenders based on other observable traits after *Booker*, such as educational attainment and criminal history. I also find evidence that the racial sentencing gap expands after periods of more deferential appellate review, suggesting that judges are particularly responsive to changes in the likelihood of appellate reversal.

Next, I examine the sources of increasing disparities after *Booker* by studying how different types of judges respond to increased judicial discretion. Many scholars have suggested that judges have different sentencing philosophies (e.g., Hofer, Blackwell, and Ruback 1999), which may be affected by the standard of appellate review (Fischman and Schanzenbach 2011), with correlations between sentencing practices and judicial characteristics such as race, gender, and political affiliation (Welch 1988, Schanzenbach 2005, Schanzenbach and Tiller 2007, Schanzenbach and Tiller 2008). However, prior empirical research on inter-judge disparity and the impact of judicial demographics on sentencing practices has been hampered by the lack of judge identifiers. Relying on aggregate district-level variation in judicial demographics can lead to biased estimates if districts with different judicial compositions differ in ways that affect all judges within the district court.

I surmount these issues by utilizing a novel dataset constructed for this study. Matching three data sources, I construct a dataset of over 400,000 criminal defendants linked to sentencing judge from fiscal years 2000-2009. Given that cases are randomly assigned to judges within a district court, any difference in sentencing practices across judges can be attributable to judge differences, rather than case composition. Exploiting the random assignment of cases to judges in this dataset, I find that much of the increases in racial disparities after *Booker* are driven by post *Booker* appointed judges, even after accounting for the fact that these judges are George W. Bush appointees. My findings suggest that judges with experience sentencing under the Guidelines may have become relatively acculturated to the Guidelines regime, compared to newer judges who began their tenure in a post *Booker* regime.

I conclude by considering the impact of judicial discretion on other actors in the criminal justice system. Arrest, charge, trial and plea bargaining decisions made earlier in the process are all ripe avenues for unwarranted bias (Anwar et al. 2012, Rehavi and Starr 2012). After *Booker*, prosecutors have commented that they are far less willing to forego charging mandatory minimums when judges ultimately sentence defendants to terms far below the Guidelines recommended minimum sentence. Consistent with this story, I find evidence that increased judicial discretion via *Booker* changes the prosecutorial treatment of statutory mandatory minimums, which *Booker* left intact. Black offenders are

far more likely to be charged with mandatory minimums than similar white offenders, and after *Booker*, black defendants are significantly more likely to face statutory minimums that exceed their Guidelines minimum compared to white defendants, consistent with prosecutors attempting to rein in judicial discretion.

The paper is structured as follows. Section 2 provides a brief legal background of the Guidelines and the *Booker* decision. Section 3 describes the data and presents summary statistics. Section 4 presents a simple conceptual framework for judicial sentencing. Section 5 provides the empirical methodology. Section 6 presents results, and Section 7 concludes.

2.2. Legal Background

2.2.1. Adoption of the United States Sentencing Guidelines

For over a century prior to the adoption of the Guidelines, judges had virtually unfettered discretion to determine the lengths of sentences. A 1977 study of Virginia state district court judges revealed that while judges generally agreed on the verdict in legal cases, they applied radically different sentences (Austin and Williams 1977). A 1988 study of federal courts similarly found that white collar offenders who committed similar offenses received very different sentences depending on the court in which they were sentenced (Wheeler et al. 1988).

By the 1970s, the legal community and public expressed alarm at the widespread disparities in criminal sentencing that resulted from this indeterminate sentencing regime (Frankel 1973). Some members of the public argued that judges and parole boards endangered public safety with lenient sentencing of criminals (Tonry 2005). Others were distressed by inequitable and arbitrary treatment within the criminal justice system. The American Friends Service Committee claimed that decreasing discretion among judges was the only way to eliminate racial discrimination in the criminal justice system (American Friends Service Committee 1971).

Policymakers also recognized that judges were often “left to apply [their] own notions of the purposes of sentencing,” leading to “an unjustifiably wide range of sentences to offenders convicted for similar crimes” (S. Rep. No. 98-225 1983). In order to eliminate unwarranted sentencing disparities “among defendants with similar records who have been found guilty of similar criminal conduct,” Congress created the United States Sentencing Commission to adopt and administer the Guidelines. Part of the SRA of 1984, the Guidelines were applied to all federal offenses committed after November 1, 1987, and prohibited courts from using race, sex, national origin, creed, religion, and socioeconomic status in sentencing decisions.

Under the Guidelines, federal district court judges assign each crime to one of 43 offense levels, and assign each defendant to one of six criminal history categories. The more serious the offense and the greater the harm associated with the offense, the higher the base offense level. For example, trespass offenses are assigned a base offense level of four, while kidnapping is assigned a base offense level of 32. From a base offense level, the final offense level is

calculated by adjusting for applicable offense and defendant characteristics. Relevant adjustments under Chapter Two of the Guidelines include the amount of loss involved in the offense, use of a firearm, and the age or condition of the victim. Chapter Three allows for further adjustments based on aggravating or mitigating factors, such as obstruction of justice or a defendant's acceptance of responsibility.

The criminal history category reflects the frequency and severity of a defendant's prior criminal convictions, predictive of recidivism risk. To determine a defendant's criminal history category, a judge adds points for prior sentences in the federal system, 50 state systems, all territories and foreign or military courts. For example, three points are added for each prior sentence of imprisonment exceeding one year and one month, and two points are added for each prior sentence of imprisonment of at least 60 days and less than one year and one month. Two points are also added if the defendant committed the instant offense under any criminal justice sentence. These points are then converted into a criminal history category.

The intersection of the final offense level and criminal history category yields a fairly narrow Guidelines recommended sentencing range, where the top of the range exceeds the bottom by the greater of either six months or 25%. See Table 2.1 for the Guidelines sentencing chart.

TABLE 2.1. GUIDELINES SENTENCING CHART

	Offense Level	Criminal History Category (Criminal History Points)					
		I (0 or 1)	II (2 or 3)	III (4, 5, 6)	IV (7, 8, 9)	V (10, 11, 12)	VI (13 or more)
Zone A	1	0-6	0-6	0-6	0-6	0-6	0-6
	2	0-6	0-6	0-6	0-6	0-6	1-7
	3	0-6	0-6	0-6	0-6	2-8	3-9
	4	0-6	0-6	0-6	2-8	4-10	6-12
	5	0-6	0-6	1-7	4-10	6-12	9-15
	6	0-6	1-7	2-8	6-12	9-15	12-18
Zone B	7	0-6	2-8	4-10	8-14	12-18	15-21
	8	0-6	4-10	6-12	10-16	15-21	18-24
	9	4-10	6-12	8-14	12-18	18-24	21-27
	10	6-12	8-14	10-16	15-21	21-27	24-30
Zone C	11	8-14	10-16	12-18	18-24	24-30	27-33
	12	10-16	12-18	15-21	21-27	27-33	30-37
	13	12-18	15-21	18-24	24-30	30-37	33-41
Zone D	14	15-21	18-24	21-27	27-33	33-41	37-46
	15	18-24	21-27	24-30	30-37	37-46	41-51
	16	21-27	24-30	27-33	33-41	41-51	46-57
	17	24-30	27-33	30-37	37-46	46-57	51-63
	18	27-33	30-37	33-41	41-51	51-63	57-71
	19	30-37	33-41	37-46	46-57	57-71	63-78
	20	33-41	37-46	41-51	51-63	63-78	70-87
	21	37-46	41-51	46-57	57-71	70-87	77-96
	22	41-51	46-57	51-63	63-78	77-96	84-105
	23	46-57	51-63	57-71	70-87	84-105	92-115
	24	51-63	57-71	63-78	77-96	92-115	100-125
	25	57-71	63-78	70-87	84-105	100-125	110-137
	26	63-78	70-87	78-97	92-115	110-137	120-150
	27	70-87	78-97	87-108	100-125	120-150	130-162
	28	78-97	87-108	97-121	110-137	130-162	140-175
	29	87-108	97-121	108-135	121-151	140-175	151-188
	30	97-121	108-135	121-151	135-168	151-188	168-210
	31	108-135	121-151	135-168	151-188	168-210	188-235
	32	121-151	135-168	151-188	168-210	188-235	210-262
	33	135-168	151-188	168-210	188-235	210-262	235-293
	34	151-188	168-210	188-235	210-262	235-293	262-327
	35	168-210	188-235	210-262	235-293	262-327	292-365
	36	188-235	210-262	235-293	262-327	292-365	324-405
	37	210-262	235-293	262-327	292-365	324-405	360-life
	38	235-293	262-327	292-365	324-405	360-life	360-life
	39	262-327	292-365	324-405	360-life	360-life	360-life
	40	292-365	324-405	360-life	360-life	360-life	360-life
	41	324-405	360-life	360-life	360-life	360-life	360-life
	42	360-life	360-life	360-life	360-life	360-life	360-life
	43	life	life	life	life	life	life

Notes: Recommended sentence lengths in months.

If a judge determines that there are aggravating or mitigating circumstances that warrant a departure from the Guidelines, she would have to justify her reasons for departure to the appellate court, but in general the Guidelines were treated as sufficiently mandatory prior to *Booker*. Before *Booker*, judges could only consider factors such as a defendant's age, education, employment history, in deciding the sentence length for within range sentences. After

sentencing, the government is permitted to appeal a sentence resulting in a departure below the Guidelines range, and the defendant can appeal an upward departure.²

2.2.2. Challenges to the Mandatory Guidelines Regime

The constitutionality of mandatory sentencing guidelines was first questioned in reference to the Washington State Sentencing Guidelines. In *Blakely v. Washington*, 542 U.S. 296 (2004), the Supreme Court held that the Sixth Amendment right to a jury trial prohibited judges from increasing a defendant's sentence beyond the statutory maximum based on facts other than those decided by the jury beyond a reasonable doubt. As a result, Washington's mandatory sentencing guidelines were struck down. Shortly after, the reasoning of *Blakely* was applied to the United States Sentencing Guidelines.

In *United States v. Booker*, the mandatory Guidelines were also found unconstitutional under the Sixth Amendment. The *Booker* ruling, however, did not apply to mandatory minimum sentences enacted by Congress. Rather than invalidating the Guidelines, the Supreme Court held that the Guidelines would be "effectively advisory," as opposed to mandatory. The Court explained that "district courts, while not bound to apply the Guidelines, must consult those Guidelines and take them into account when sentencing."

In the aftermath of *Booker*, circuit courts reached a consensus that sentencing must begin with the calculation of the applicable Guidelines range. Today, after a sentencing judge has calculated the Guidelines range, she must consider seven factors under 18 U.S.C. §3553(a) before imposition of punishment: 1) the nature and circumstances of the offense and the history and characteristics of the defendant, 2) the need for the sentence imposed, 3) the kinds of sentences available, 4) the kinds of sentence and the sentencing range established, 5) any pertinent policy statement issued by the Sentencing Commission, 6) the need to avoid unwarranted sentence disparities among defendants with similar records who have been found guilty of similar conduct, and 7) the need to provide restitution to any victims of the offense.

Subsequent Supreme Court decisions furthered weakened the effect of the Guidelines on criminal sentencing by reducing the degree of appellate review. In *Rita v. United States*, 551 U.S. 338 (2007), the Court held that a sentence within the Guidelines recommended range could be presumed "reasonable" because a "judge who imposes a sentence within the range recommended by the Guidelines thus makes a decision that is fully consistent with the Commission's judgment in general." In *Gall v. United States*, 552 U.S. 38 (2007), the Court held that federal appeals courts could not presume that a sentence outside the range recommended by the Guidelines was unreasonable. Concurrent with the *Gall* decision, the Court in *Kimbrough v. United States*, 552 U.S. 85 (2007), held that federal district court judges

²There are numerous other ways in which Congress has attempted to limit unwarranted disparities in sentencing. Beginning in 1984, and subsequently 1986 and 1988, Congress enacted a series of mandatory minimum statutes directed at drug and firearms offenses. In 2003, Congress also passed the PROTECT Act to curtail judicial departures due to a concern that the standard for appellate review of departures had led to undesirably high rates of departures for child sex offenses.

have the discretion to impose sentences outside the recommended Guidelines range due to policy disagreements with the Sentencing Commission, such as the disparate treatment of crack and powder cocaine offenses - the so-called “100-to-1 ratio.”

2.3. Data

This paper utilizes data from three sources: 1) the United States Sentencing Commission, 2) the Transactional Records Access Clearinghouse, and 3) the Federal Judicial Center. I describe each dataset in turn.

2.3.1. United States Sentencing Commission

I use data from the United States Sentencing Commission (USSC) on records of all federal offenders sentenced pursuant to the Sentencing Guidelines and Policy Statements of the SRA of 1984 in fiscal years 1994-2009 (October 1, 1993 - September 30, 2009).³ These data include demographic, Guidelines application, and sentencing information on federal defendants, but defendant and judge identifiers are redacted. This information is obtained from numerous documents on every offender: Indictment, Presentence Report, Report on the Sentencing Hearing, Written Plea Agreement (if applicable), and Judgment of Conviction.

Demographic variables include defendant’s race, gender, age, citizenship status, educational attainment, and number of dependents. Data is also provided on the primary offense type, with a total of 35 offense categories. Offense level variables include the base offense level, the base offense level after Chapter Two adjustments and the final offense level after Chapter Three adjustments. Criminal history variables include whether the defendant has a prior criminal record, and whether armed career criminal status, or career offender status is applied, which are subject to mandatory minimums. Data is also provided on the total number of criminal history points applied and the final criminal history category.

For each offender, there is a computed Guidelines range, and a Guidelines range adjusted for applicable mandatory minimums. From these variables, I construct indicator variables for above range and below range departures from the Guidelines, as well as months of departure, conditional on an above or below range departure.⁴ Information is also provided on whether the offense carries a mandatory minimum sentence under various statutes, and whether departures from the statutory minimum are granted, either under a substantial assistance motion or application of the safety valve (described in greater detail later). Sentencing characteristics include the district court in which sentencing occurred (94 total), in addition to the sentencing month and year.⁵ Data is also available on whether a case is settled by plea

³Over 90% of felony defendants in the federal criminal justice system are sentenced pursuant to the SRA of 1984 and all cases are assessed to be constitutional.

⁴Technically, deviations from the Guidelines range are no longer “departures” after the Guidelines became advisory, but I use this term to maintain consistency.

⁵USSC data prior to 2004 actually includes information on the exact sentencing day, but this variable is not available in later years.

agreement or trial. Sentencing outcomes include incarceration or probation, sentence length, receipt of supervised release, and length of supervised release.

I apply several sample restrictions. First, I drop individuals sentenced to life imprisonment, about 0.5% of the sample. Second, I drop individuals with missing or invalid criminal records (offense level, criminal history category, and offense type), about 6% of the sample. Third, I exclude individuals missing race, about 0.2% of the sample.

Table 2.2 presents summary statistics for the main variables from the USSC data. Panel A indicates that 83% of the defendants in the dataset are incarcerated versus receiving probation. Those who are not incarcerated serve an average of 29 months of probation. The average unconditional sentence length is approximately 49 months. Conditional on incarceration, the average sentence length is 57 months. Approximately 30% of cases carry a statutory minimum and only 4% of cases are settled by trial. After imprisonment, defendants serve an average of 38 months of supervised release.

In the dataset, 32% of defendants are white, 26% black and 38% Hispanic.⁶ About 32% of the defendants are non U.S. citizens. Defendants have on average 1.6 dependents, and almost a majority have less than a high school degree. Over 85% of the defendants are male. Defendants are approximately 34 years of age. Most of the defendants have had some previous interaction with the criminal justice system, as 75% have some prior criminal history. The most common offense is drug trafficking, followed by immigration, fraud, firearms, and larceny. Drug trafficking represents about 39% of the cases, followed by immigration offenses which comprise 18% of the cases. In terms of Guidelines range calculations, defendants have an average final criminal history score of 2.36, and a final offense level of 18.84. This criminal history category and offense level combination yield an average Guidelines recommended range of 30-37 months in prison.

⁶The remaining race category is defendants classified as “other” race, which is comprised primarily of Native Americans.

TABLE 2.2. SUMMARY STATISTICS

PANEL A. USSC DATA, 1994-2009					
Variable	Obs	Mean	Std. Dev.	Min	Max
Incarceration	853008	0.833	0.373	0	1
Probation Length in Months	142627	29.858	22.238	0	997
Sentence Length in Months	847227	49.290	65.108	0	985
Statutory Minimum Applied	853561	0.299	0.458	0	1
Settled by Trial	665073	0.044	0.205	0	1
Supervised Release in Months	852701	38.490	59.829	0	999
White	852875	0.318	0.466	0	1
Black	852875	0.261	0.439	0	1
Hispanic	852875	0.379	0.485	0	1
Non US Citizen	852990	0.320	0.467	0	1
Number of Dependents	854992	1.598	2.023	0	98
Less Than High School	842099	0.461	0.498	0	1
Male	854611	0.859	0.348	0	1
Age	854992	34.696	10.798	16	98
Criminal History Indicator	664422	0.746	0.435	0	1
Drug Trafficking Offense	854992	0.388	0.487	0	1
Immigration Offense	854992	0.179	0.383	0	1
Fraud Offense	854992	0.113	0.317	0	1
Firearm Offense	854992	0.092	0.289	0	1
Criminal History Category (1-6)	854992	2.361	1.699	1	6
Final Offense Level (1-43)	854992	18.841	8.961	1	43

PANEL B. JUDGE MATCHED DATA, 2000-2009					
Variable	Obs	Mean	Std. Dev.	Min	Max
Incarceration	643990	0.839	0.368	0	1
Probation Length in Months	103822	25.345	22.065	0	120
Sentence Length in Months	641986	45.920	59.871	0	985
Statutory Minimum Applied	633235	0.282	0.450	0	1
Settled by Trial	643990	0.035	0.183	0	1
Supervised Release in Months	643347	38.264	61.330	0	999
White	626500	0.294	0.456	0	1
Black	626500	0.234	0.423	0	1
Hispanic	626500	0.436	0.496	0	1
Non US Citizen	633942	0.384	0.486	0	1
Number of Dependents	595781	1.616	1.739	0	82
Less Than High School	599619	0.489	0.499	0	1
Male	636641	0.867	0.340	0	1
Age	638530	34.548	10.644	16	97
Criminal History Indicator	632772	0.749	0.434	0	1
Drug Trafficking Offense	643990	0.369	0.482	0	1
Immigration Offense	643990	0.251	0.433	0	1
Fraud Offense	643990	0.101	0.301	0	1
Firearm Offense	643990	0.096	0.295	0	1
Criminal History Category (1-6)	643990	2.416	1.705	1	6
Final Offense Level (1-43)	643990	18.451	8.625	1	43
Male Judge	643990	0.807	.395	0	1
White Judge	643990	0.767	.423	0	1
Black Judge	643990	0.083	.275	0	1
Hispanic Judge	643990	0.140	.347	0	1
Democratic Judge	643990	0.437	.496	0	1

Notes: Panel A is from the USSC data from 1994-2009. Panel B is from the USSC, TRAC, and Federal Judicial Center matched data from 2000-2009.

2.3.2. Transactional Records Access Clearinghouse

The Transactional Records Access Clearinghouse (TRAC) provides sentencing data obtained through FOIA requests. The data do not contain defendant demographics, offense characteristics, and Guidelines application informa-

tion, but defendants are linked to the sentencing judge. To link defendant and crime characteristics to sentencing judge, I match sentencing records from the USSC to data provided by TRAC. By district court, matching is conducted on several key variables: sentencing year, sentencing month, offense type, sentence length in months, probation length in months, amount of monetary fine, whether the case ended by trial or plea agreement, and whether the case resulted in a life sentence. For defendants sentenced prior to fiscal year 2004, I also match on exact sentencing day.⁷ I successfully match over 90% of all cases from fiscal years 2000-2009.

2.3.3. Federal Judicial Center

To provide information on judge characteristics, I match the USSC and TRAC combined data to judge biographical data from the Federal Judicial Center. Federal district judges are Article III judges who serve life term tenures. New appointments are generally made when a judge retires or dies.⁸ As of the current day, there are a total of 678 Article III district judgeships. The largest district court is the Southern District of New York with 28 authorized judgeships. The majority of other district courts have between two to seven judgeships.

I obtain information on judge race, gender, political affiliation of appointing president, and commission year. Applying the same sample restrictions as described in Section 3.1, the final matched dataset consists of 440,025 cases resulting in prison sentences from fiscal years 2000-2009.⁹ This unique dataset permits an examination of judicial demographic characteristics on sentencing practices in the wake of increased judicial discretion via *Booker*. Panel B of Table 1 presents summary statistics on this matched dataset. Of judges active between 2000-2009, 19% are female, and over 75% are white. Black judges represent approximately 8% of the share of all judges. Judges appointed by Democratic presidents represent 44% of all judges.

2.4. Conceptual Framework of Judicial Sentencing

This section provides a very simplified framework for analyzing judicial sentencing, similar to that used by Genaioli and Shleifer (2008). The framework considers two countervailing forces on judicial sentencing: a judge's preferences for sentencing according to her tastes and costs associated with exercising discretion.

Consider a judge who is assigned to a defendant with a true harm or risk of recidivism, r . Let the Guidelines sentence for a defendant with risk r be $s^*(r)$. Now suppose that the judge would prefer to sentence the defendant to $s^j(r)$, such that $s^j(r) \neq s^*(r)$. The judge may prefer to impose $s^j(r)$ because sentencing a defendant in a particular way can increase the judge's utility by advancing her political and ideological goals, or other personal goals.¹⁰

⁷Results are unchanged matching on the same variables across all years.

⁸On a few occasions, Congress has also increased the number of judgeships within a district in response to changing population or caseload.

⁹The Federal Judicial Center does not collect demographic information on judges in 3 districts: Guam, Virgin Islands, and Northern Mariana Islands.

¹⁰65% of federal district judges in a 2010 USSC survey indicated that they thought the departure policy statements in the Guidelines Manual were too restrictive, indicating that many judges prefer to deviate from the Guidelines.

Assume that a judge suffers disutility from sentencing $s \neq s^j(r)$, such that a judge who sentences s experiences loss of

$$L = \frac{(s - s^j(r))^2}{2}$$

If judges have sentencing preferences that deviate from the Guidelines and were left unconstrained, a judge would set $s = s^j(r)$ and one would likely observe large inter-judge disparities in sentencing. Consistent with this prediction, Posner (2005) suggests that the large variances in federal sentences prior to the adoption of the Guidelines were likely due to differing judicial attitudes towards personal responsibility and deterrence.

However, various mechanisms constrain judges from deviating from recommended sentences. For one, mandatory rule-based sentencing under a Guidelines regime constrains judge sentencing. Another constraint on judge decision-making comes from appellate review. A high reversal rate is not only administratively burdensome, but also potentially harms a trial judge's prospects for promotion to the appeals court (Posner 2005).

Thus, a judge sentencing away from the Guidelines recommended sentence also incurs a cost associated with reversal. Assume that pre *Booker*, a judge faced a cost $C = 0$ if $s = s^*(r)$ and $C = \infty$ if $s \neq s^*(r)$. Essentially, the Guidelines were treated as mandatory, implying very high costs to exercising discretion. In this pre *Booker* regime, one would see very little deviation from the Guidelines.¹¹

After *Booker*, the Guidelines were no longer binding, but judges still faced the prospect of reversal upon appellate review. To capture this idea, assume that the cost of exercising discretion in the post *Booker* regime is

$$C = p \frac{(s^j(r) - s^*(r))^2}{2}$$

where p is the probability of appellate reversal and $\frac{(s^j(r) - s^*(r))^2}{2}$ is the reputational cost associated with reversal.

Given a defendant with true risk r , a judge therefore sets a sentence $s(r)$ to minimize $\frac{(s - s^j(r))^2}{2} + p \frac{(s^j(r) - s^*(r))^2}{2}$, setting

$$s(r) = \frac{s^j(r) + ps^*(r)}{1 + p}$$

Thus, the judge imposes a sentence that is a weighted average of his ideal and the Guidelines recommended sentence. If $p = 0$, he sets the sentence to his ideal. The greater the probability of reversal, p , the more the judge sentences the defendant closer to the Guidelines sentence.

From a Guidelines regime to *Booker*, the total cost of exercising discretion C falls substantially for judges who want to depart from the Guidelines sentence. *Rita*, *Gall*, and *Kimbrough* later reduced the level of appellate review

¹¹The rate of departure from the Guidelines was less than 15% in the early 1990s.

from *de novo* to substantial abuse of discretion, intuitively lowering p , the probability of appellate reversal. Indeed, the probability of reversal on sentencing matters fell from 36% in 2006 (under *de novo* review), to 26% in 2008 (under abuse of discretion review).¹² Thus, as the cost of exercising discretion falls after *Booker*, the model predicts that judges would immediately impose sentences that deviate from the Guidelines sentence. As the probability of appellate reversal falls under *Rita*, *Gall*, *Kimbrough*, the costs of discretion fall even more, and one would expect to see further deviations from the Guidelines. If the probability of appellate reversal under *de novo* review was a binding constraint on judges, one would expect to see relatively larger changes in sentencing after *Rita*, *Gall*, and *Kimbrough* than in the immediate aftermath of *Booker*.

2.5. Empirical Methodology

The *Booker* case was decided on January 12, 2005, and applied immediately to all future cases and prior cases that had not reached sentencing. This paper exploits the timing of this decision to estimate the effect of increased judicial discretion on racial disparities in sentencing outcomes. I use a differences-in-differences (DD) methodology to compare the sentence disparities between similar defendants within a district court before and after *Booker*.

The main specification is of the form:

$$Y_{ijkdtm} = \beta_0 + \beta_1 * Booker * Race_i + \beta_2 * \mathbf{X}_i + \beta_3 * Z_i + Guide_{ijk} + Offtype_i + \gamma_d + \delta_t + \gamma_d * \delta_t + \lambda_m + \epsilon_{ijkdtm} \quad (1)$$

where Y_{ijkdtm} is a sentencing outcome for defendant i , with criminal history category j and offense level k , sentenced in district court d in year t and month m . Main outcomes include sentence length measured in months, a binary indicator for whether the defendant was sentenced above range (such that the sentence length is greater than the prescribed Guidelines maximum), a binary indicator for below range sentencing (sentence length less than the prescribed Guidelines minimum), and sentence length conditional on above, below, or within range sentencing. Additional outcomes include a binary indicator for incarceration, probation length, receipt of supervised release, term of supervised release, application of a statutory minimum, and departures from statutory minimums.

The main coefficient of interest is β_1 , which captures the impact of *Booker* on racial gaps in sentencing outcomes. *Booker* is an indicator variable for defendants sentenced after the *Booker* decision.¹³ $Race_i$ is a dummy variable for defendant i 's race: white, black, Hispanic, or other. The *Booker* indicator and main race dummies are also included in the specification. \mathbf{X}_i comprises a vector of demographic characteristics of the defendant including gender, age, age squared, educational attainment (less than high school, high school graduate, some college, college graduate), number

¹²I calculate rate of appellate reversals using yearly data on the universe of criminal appeals from the USSC. Reversal is defined as all reversals and remands on appeals arising out of sentencing issues.

¹³For defendants sentenced in January 2005, the USSC data contains a variable denoting whether the case was heard prior to or after the *Booker* decision.

of dependents, and citizenship status. Z_i , an indicator variable for whether the offense carries a mandatory minimum.

$Guide_{ijk}$ includes dummy variables for criminal history category j and offense level k , and each unique combination of criminal history category and offense level. The interaction captures differential sentencing tendencies at each unique cell of the Guidelines grid (258 total). To proxy for underlying offense seriousness and all aggravating and mitigating factors, I control for final offense level. I also control for final criminal history category. $Offtype_i$ is a dummy variable for offense type.

The specification also includes district court fixed effects (γ_d), sentencing year fixed effects (δ_t), and sentencing month fixed effects (λ_m). I also control for district by year fixed effects to control for district trends over time. As a robustness check, race specific linear trends are included to account for preexisting trending differences in sentencing outcomes between defendants of different races. All standard errors are clustered at the district court level to account for serial correlation.

To analyze the differential sentencing practices of certain types of judges, I use a differences-in-differences-in-differences (DDD) methodology. The DDD methodology captures how judges differ in their relative treatment of similar black and white defendants in response to increased judicial discretion, compared to other judges within the same district court. Because cases are randomly assigned to judges within a district court, judge identifiers allow one to compare judges within the same court, capturing judge differences in sentencing rather than different caseloads.¹⁴

I identify the sources of increasing racial disparities post *Booker* using a specification of the form:

$$Y_{ijkdtm} = \beta_0 + \alpha * Judge_i * Race_i * Booker + \beta_1 * Booker * Race_i + \beta_2 * Judge_i * Race_i + \beta_3 * \mathbf{X}_i + \beta_4 * Z_i + Guide_{ijk} + Offtype_i + \gamma_d + \delta_t + \gamma_d * \delta_t + \lambda_m + \epsilon_{ijkdtm} \quad (2)$$

where $Judge_i$ includes judicial demographics such as race, gender, political affiliation, an indicator for pre vs. post Guidelines appointment, and an indicator for pre vs. post *Booker* appointment. The coefficient α captures the impact of particular judicial characteristics on racial disparities in the wake of *Booker*.

2.5.1. Potential Threats to Identification

The results presented in this paper may be biased if unobservables that affect sentencing decisions change differentially by defendant race in the wake of *Booker*. I test for this potential concern by analyzing the extent to which observable offense and defendant characteristics differ in the post *Booker* period. I find that there is no differential change in criminal history by defendant race after *Booker*. If anything, black defendants have lower base offense levels and lower final offense levels after *Booker* compared to their white counterparts, suggesting that black defendants

¹⁴According to the Administrative Office of the United States Courts, “The majority of courts use some variation of a random drawing.” I also test for random assignment in Section 6.5.

commit relatively less severe crimes compared to similar white offenders (See Table 2.3).

TABLE 2.3. DEFENDANT CRIMINAL CHARACTERISTICS

	(1) Criminal History	(2) Total Criminal Points	(3) Criminal History Category	(4) Base Offense Level	(5) Final Offense Level
Booker*Black	0.00347 (0.00623)	0.104 (0.0984)	0.0330 (0.0261)	-0.356*** (0.135)	-0.413*** (0.112)
Booker*Hispanic	-0.00669 (0.00769)	0.0438 (0.252)	-0.0295 (0.0352)	-0.0805 (0.186)	-0.835*** (0.148)
Booker*Other	0.0277** (0.0133)	0.115 (0.156)	0.0208 (0.0517)	0.431 (0.298)	0.228 (0.246)
Black	0.0800*** (0.00583)	1.717*** (0.104)	0.607*** (0.0269)	0.604*** (0.147)	0.858*** (0.153)
Hispanic	-0.0236*** (0.00787)	-0.445** (0.177)	-0.0753** (0.0309)	0.451 (0.278)	0.703*** (0.261)
Other	-0.0671*** (0.0146)	-0.980*** (0.150)	-0.273*** (0.0535)	-0.330 (0.264)	-0.0818 (0.231)
Booker	-0.00223 (0.0127)	-0.224 (0.185)	-0.0398 (0.0366)	0.756*** (0.196)	1.239*** (0.203)
Race Trends?	Yes	Yes	Yes	Yes	Yes
Observations	636,698	822,908	824,680	553,759	824,680
R-squared	0.224	0.206	0.307	0.541	0.522

Notes: Data is from the USSC from 1994-2009. When the dependent variable is the base offense level, data is from 1999-2009. Regressions for criminal history, total criminal history points and criminal history category contain controls for final offense level. Regressions for base and final offense level control for criminal history category. All regressions contain controls for offense type. Regressions also contain district by sentencing year, and sentencing month fixed effects, and standard errors are clustered at the district level. Race trends are included. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Black defendants sentenced after *Booker* are more likely to be male, 0.7 years younger, and less likely to be non U.S. citizens compared to their white counterparts (See Table 2.4). While these changes are significant, the magnitudes are very small. Moreover, as shown later, younger defendants who are U.S. citizens receive relatively lower sentences compared to otherwise similar, older non U.S. citizens. Thus, any unobservable changes correlated with these demographics would bias downwards my findings. Overall, these results suggests that unobservables are unlikely to change in large enough ways to upwards bias my estimates.

TABLE 2.4. DEFENDANT DEMOGRAPHIC CHARACTERISTICS

	(1) Male	(2) Age	(3) Number Dependents	(4) Non US Citizen	(5) Less than HS
Booker*Black	0.0234*** (0.00698)	-0.715*** (0.139)	0.0375 (0.0246)	-0.0209*** (0.00734)	-0.00324 (0.00553)
Booker*Hispanic	0.00756 (0.00531)	-0.113 (0.156)	0.0335 (0.0312)	0.00909 (0.00980)	0.00125 (0.00709)
Booker*Other	0.00976 (0.0110)	-0.568 (0.342)	-0.0553 (0.0517)	0.0205* (0.0108)	0.0247** (0.0117)
Black	-0.0258** (0.0100)	-4.455*** (0.166)	0.595*** (0.0287)	0.0274** (0.0120)	0.0743*** (0.00768)
Hispanic	0.00317 (0.00665)	-4.086*** (0.194)	0.463*** (0.0301)	0.408*** (0.0183)	0.213*** (0.0116)
Other	-0.0354*** (0.00803)	-3.227*** (0.271)	0.267*** (0.0521)	0.162*** (0.0314)	0.0319* (0.0178)
Booker	-0.00310 (0.00822)	0.400 (0.278)	-0.0674 (0.0473)	-0.00134 (0.00978)	-0.0138 (0.0128)
Race Trends?	Yes	Yes	Yes	Yes	Yes
Observations	824,680	824,680	824,680	824,680	824,680
R-squared	0.135	0.182	0.125	0.568	0.240

Notes: Data is from the USSC from 1994-2009. All regressions contain controls for offense type, and dummies for each offense level and criminal history combination. Regressions also contain district by sentencing year, and sentencing month fixed effects, and standard errors are clustered at the district level. Race trends are included. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Another potential threat to the identification is if *Booker* is associated with changes in selection in the types of defendants that reach the sentencing stage. For instance, if prosecutors disproportionately drop or dismiss charges against marginal black defendants, the remaining black defendants at the sentencing stage might receive longer sentences compared to similar white offenders. To address potential changes in selection prior to the sentencing stage, I test the likelihood of guilty pleas, dropped charges, and deferred prosecutions against black defendants compared to similar white defendants after *Booker* using data on all federal arrests and bookings from 1994-2009.¹⁵ Table 2.5 suggests no significant changes in the rates at which black defendants plead guilty, or the likelihood of dropped charges or deferred prosecution, suggesting no significant changes in selection prior to sentencing.¹⁶

¹⁵Data is obtained from the Federal Justice Statistics Program: Arrests and Bookings for Federal Offenses, which covers all offenders within the custody of the United Marshals Service.

¹⁶A deferred prosecution occurs when a prosecutor agrees to not file charges in exchange for the defendant taking specified actions, such as payment of fines, and continued cooperation during investigation.

TABLE 2.5. SELECTION INTO SENTENCING STAGE

	(1) Guilty Plea	(2) Dropped Charge	(3) Deferred Prosecution
Booker*Black	-0.0115 (0.00808)	-0.00451* (0.00242)	0.000103 (0.000901)
Black	-0.00718 (0.00948)	0.00634*** (0.00218)	-0.000823 (0.000854)
Booker	-0.112*** (0.00779)	-0.00186 (0.00351)	-5.74e-05 (0.000385)
Observations	1,669,560	1,669,560	1,669,560
R-squared	0.241	0.043	0.032

Notes: Data is from the Arrests and Bookings for Federal Offenses from 1994-2009. All regressions contain controls for defendant gender, age, marital status, citizenship status, primary offense type, district court by arrest year fixed effects, and race trends. Standard errors are clustered at the district level. The coefficient of interest is the interaction of defendant race (omitted group white defendants) with a *Booker* indicator for defendants arrested after *Booker*. Race trends are included. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

2.6. Results

2.6.1. Sentence Length

Figure 2.1 presents graphical evidence of trends in sentence length by defendant race in the raw data. Figure 2.1 indicates no preexisting trending differences in sentence lengths across defendants of different races. However, the trend in the gap in sentence length between black and white defendants changes post *Booker* as sentence lengths for black and white defendants diverge. The evidence is even more striking excluding cases with mandatory minimums, where it is apparent that sentence lengths for white defendants decrease post *Booker*, while black sentence lengths continue to rise, increasing the racial disparities in sentence length.

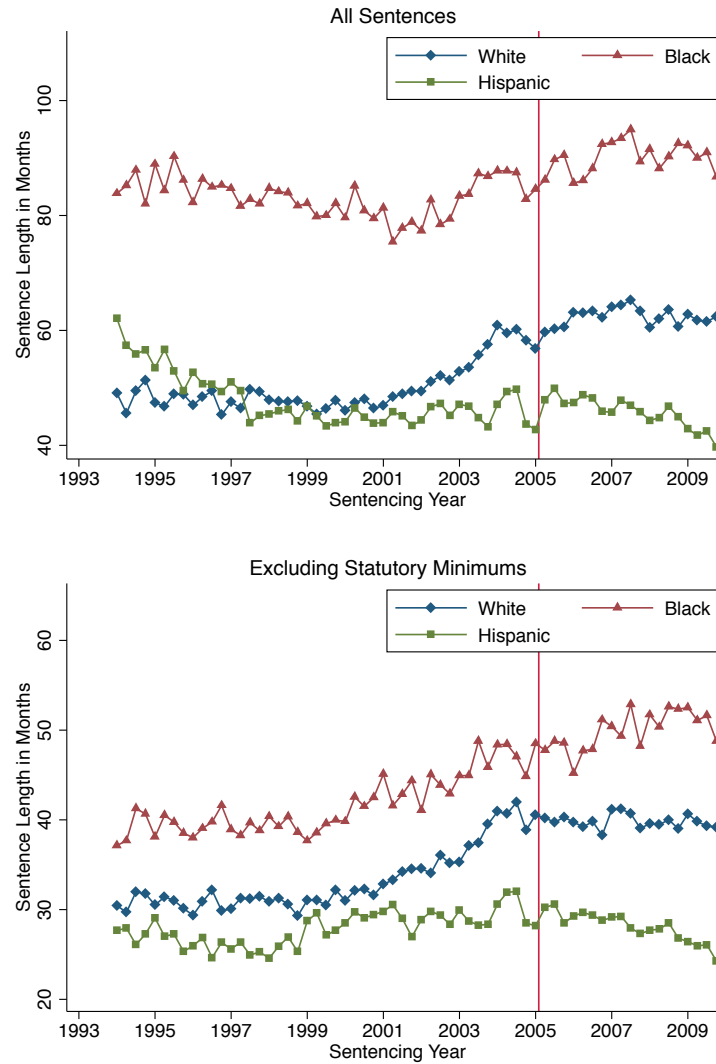


FIGURE 2.1. SENTENCE LENGTHS IN MONTHS, BY DEFENDANT RACE

Notes: Data is from the USSC from 1994-2009. Data points are quarterly averages.

Table 2.6 presents the regression results for the impact of increased judicial discretion via *Booker* on disparities in sentence length. The coefficients on defendant demographics are consistent with prior findings regarding disparities in sentencing. On average, black offenders face an approximately 3 month longer sentence length than comparable white offenders, who are the omitted category. Hispanic offenders receive over a 1 month longer prison sentence compared to similar white offenders. Additionally, non US citizens face about a 1.5 month longer prison sentence compared to US citizens. Defendants with greater educational attainment receive shorter months in prison, compared to defendants with less than a high school degree (the omitted category). I also find large sentencing disparities by gender. Female defendants receive over 5 months less in prison compared to male offenders. Additionally, defendant age is positively correlated with sentence length, while number of dependents is negatively associated with sentence

length. The application of a mandatory minimum on average results in a 23 month longer sentence.

The coefficients on the *Booker* indicator interacted with defendant race suggest growing racial disparities post *Booker*. Column 1 suggests that black offenders receive an approximately 2 month longer sentence after *Booker* compared to white offenders, over a 70% increase in the racial gap in sentence length, and a 4% increase in the average sentence length for all offenders. Post *Booker*, Hispanics offenders receive about a 1.5 month longer sentence compared to similar white offenders, an approximately 3% increase in the average sentence length for all offenders.

TABLE 2.6. SENTENCE LENGTH IN MONTHS

	(1)	(2)	(3)	(4)	(5)
	Sentence	Sentence	Sentence	Sentence	Sentence
Booker*Black	2.373*** (0.595)	1.653*** (0.524)	1.343** (0.617)	1.639** (0.690)	1.642** (0.689)
Booker*Hispanic	1.687*** (0.446)	1.559*** (0.547)	1.112** (0.499)	1.098** (0.539)	1.113** (0.539)
Booker*Other	2.711*** (0.986)	2.053** (1.021)	2.295** (1.113)	-0.168 (1.235)	-0.105 (1.229)
Black	2.638*** (0.363)	2.485*** (0.411)	2.639*** (0.363)	3.185*** (0.591)	3.188*** (0.591)
Hispanic	0.878* (0.461)	0.850* (0.463)	0.877* (0.461)	1.308** (0.524)	1.306** (0.522)
Other	1.061 (1.092)	0.903 (1.271)	1.057 (1.092)	3.177*** (1.055)	3.142*** (1.050)
Booker	-3.144*** (1.010)	-2.840*** (1.077)	-2.656*** (0.995)	-2.593** (1.129)	-3.671*** (1.323)
Non US Citizen	1.466*** (0.450)	1.470*** (0.452)	1.472*** (0.450)	1.479*** (0.452)	1.478*** (0.450)
HS Grad	-0.554*** (0.185)	-0.554*** (0.185)	-0.555*** (0.185)	-0.554*** (0.185)	-0.546*** (0.185)
Some College	-1.633*** (0.180)	-1.631*** (0.181)	-1.633*** (0.180)	-1.633*** (0.180)	-1.627*** (0.180)
College Grad	-1.896*** (0.235)	-1.897*** (0.235)	-1.897*** (0.235)	-1.900*** (0.235)	-1.893*** (0.235)
# Dependents	-0.150*** (0.0440)	-0.150*** (0.0440)	-0.150*** (0.0440)	-0.150*** (0.0441)	-0.150*** (0.0439)
Female	-5.388*** (0.502)	-5.387*** (0.502)	-5.385*** (0.501)	-5.386*** (0.501)	-5.385*** (0.501)
Age	0.149*** (0.0384)	0.148*** (0.0386)	0.148*** (0.0384)	0.147*** (0.0386)	0.147*** (0.0385)
Age ²	-0.00147*** (0.000429)	-0.00146*** (0.000430)	-0.00146*** (0.000429)	-0.00145*** (0.000431)	-0.00145*** (0.000429)
Mandatory Min	23.15*** (1.752)	23.15*** (1.752)	23.14*** (1.751)	23.15*** (1.752)	23.15*** (1.752)
RGK			-1.954*** (0.570)		
RGK*Black			2.112*** (0.616)		
RGK*Hispanic			1.198** (0.502)		
RGK*Other			0.871 (1.037)		
Race*PROTECT?	No	Yes	No	No	No
Race*RGK?	No	No	Yes	No	No
Race Trends?	No	No	No	Yes	Yes
Year*Month FE?	No	No	No	No	Yes
Observations	679,159	679,159	679,159	679,159	679,159
R-squared	0.741	0.741	0.741	0.741	0.741

Notes: Data is from the USSC from 1994-2009. All regressions contain controls for offense type, and dummies for each offense level and criminal history combination. Regressions also contain district by sentencing year, and sentencing month fixed effects, and standard errors are clustered at the district level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

I present several robustness checks in Table 2.6. Column 2 controls for possible differential effects of the PROTECT Act on racial disparities in sentencing outcomes. Column 3 accounts for potential effects of the 2007 *Rita/Gall/Kimbrough* decisions on racial disparities.¹⁷ Column 3 indicates that while racial disparities first emerge in the immediate aftermath of *Booker*, they grow larger following *Rita*, *Gall*, and *Kimbrough*, suggesting that judges are particularly

¹⁷I control for possible differential effects of the PROTECT Act and *Rita/Gall/Kimbrough* by interacting indicators for these court decisions with defendant race dummies. Although not shown in Table 2.6, the passage of the PROTECT Act did not change racial disparities in sentencing. This finding is also confirmed by Freeborn and Hartmann 2010.

responsive to more deferential appellate review.

Column 4 includes race specific linear trends. Column 5 includes race trends and adds a full set of time effects - sentencing month interacted with sentencing year.¹⁸

TABLE 2.7. SENTENCE LENGTH IN MONTHS - ROBUSTNESS CHECKS

	(1) Sentence	(2) Sentence	(3) Sentence	(4) Sentence	(5) Sentence
Booker*Black	2.311*** (0.683)	2.514*** (0.575)	1.355*** (0.309)	2.588*** (0.658)	2.340*** (0.595)
Booker*Hispanic	1.466*** (0.487)	1.830*** (0.450)	1.695*** (0.356)	1.681*** (0.466)	1.673*** (0.449)
Booker*Other	1.848 (1.149)	1.491 (0.932)	2.249*** (0.735)	2.927** (1.137)	2.646*** (0.983)
Black	2.885*** (0.422)	3.555*** (0.344)	0.577*** (0.190)	4.501*** (0.474)	2.680*** (0.361)
Hispanic	0.862* (0.478)	1.596*** (0.410)	-0.0824 (0.267)	0.829 (0.530)	0.863* (0.455)
Other	1.781 (1.200)	1.140 (1.036)	0.474 (0.591)	2.909*** (1.022)	1.079 (1.090)
Booker	-2.939*** (1.069)	-3.232*** (0.991)	-2.126*** (0.709)	-1.340 (1.206)	-3.103*** (1.015)
Non US Citizen	1.366*** (0.495)	1.584*** (0.454)	0.364 (0.304)	-0.176 (0.557)	1.445*** (0.449)
HS Grad	-0.369 (0.236)	-0.526*** (0.182)	-0.225** (0.0877)	-0.0568 (0.240)	-0.554*** (0.182)
Some College	-1.523*** (0.285)	-1.790*** (0.176)	-0.752*** (0.117)	-0.644** (0.301)	-1.624*** (0.178)
College Grad	-2.222*** (0.489)	-2.742*** (0.225)	-1.010*** (0.175)	0.335 (0.355)	-1.882*** (0.235)
# Dependents	-0.123 (0.136)	-0.149*** (0.0450)	-0.110*** (0.0252)	-0.106* (0.0598)	-0.140*** (0.0439)
Female	-5.347*** (0.595)	-5.502*** (0.458)	-2.633*** (0.223)	-7.139*** (0.656)	-5.416*** (0.498)
Age	0.134** (0.0622)	0.168*** (0.0394)	0.255*** (0.0255)	0.328*** (0.0647)	0.138*** (0.0377)
Age ²	-0.00125 (0.000757)	-0.00192*** (0.000443)	-0.00267*** (0.000304)	-0.00216*** (0.000674)	-0.00139*** (0.000423)
Mandatory Min	23.89*** (1.994)			34.25*** (2.135)	22.92*** (1.773)
Mandatory Min Length		0.00609*** (0.00135)			
Observations	552,524	679,159	440,930	455,203	678,960
R-squared	0.650	0.732	0.771	0.642	0.741

Notes: Data is from the USSC from 1994-2009. Column 1 presents results for all sentences including life sentences top coded at 470 months. Column 2 presents results controlling for mandatory minimum length. Column 3 presents results excluding sentences with statutory mandatory minimums. Column 4 presents results controlling for Chapter 2 adjusted offense level, which is only available for years 1999-2009. Column 5 presents results controlling for armed career criminal and career offender classification. All regressions contain controls for offense type, and dummies for each offense level and criminal history combination. Regressions also contain district by sentencing year, and sentencing month fixed effects, and standard errors are clustered at the district level. Race trends are excluded. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Finally, in Table 2.8, I replicate specification 4 for ten placebo periods prior to *Booker*. Table 2.8 indicates that the changes in racial disparities post *Booker* are much larger than those around placebo periods. Overall, these alternate specifications indicate that increases in racial disparities in the aftermath of *Booker* are highly robust.

¹⁸These results are robust to controlling for base offense level after Chapter Two adjustments and additional controls for classification as a career offender and armed career criminal (Table 2.7).

TABLE 2.8. SENTENCE LENGTH IN MONTHS - PLACEBO TESTS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Sentence 1995	Sentence 1996	Sentence 1997	Sentence 1998	Sentence 1999	Sentence 2000	Sentence 2001	Sentence 2002	Sentence 2003	Sentence 2004
Placebo Case										
Placebo*Black	-2.545** (1.080)	-2.924*** (0.921)	-2.429** (0.928)	-1.743** (0.787)	-1.907*** (0.709)	-0.961 (0.642)	-0.959 (0.649)	0.0188 (0.804)	0.672 (0.780)	0.867 (0.739)
Placebo*Hispanic	-2.373** (0.975)	-1.758* (0.941)	-0.845 (0.738)	-0.355 (0.705)	-0.838 (0.602)	-0.461 (0.602)	-0.653 (0.675)	-0.316 (0.691)	-0.415 (0.658)	0.688 (0.559)
Placebo*Other	-0.854 (2.510)	-1.739 (1.919)	-0.504 (1.899)	0.470 (1.969)	0.549 (1.623)	1.735 (1.196)	0.929 (1.826)	-0.00518 (1.974)	-0.878 (1.876)	0.261 (1.464)
Race Trends?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	692,039	692,039	692,039	692,039	692,039	692,039	692,039	692,039	692,039	692,039
R-squared	0.741	0.741	0.741	0.741	0.741	0.741	0.741	0.741	0.741	0.741

Notes: Data is from the USSC from 1994-2009. Coefficients are from DD regressions of placebo case decisions on racial disparities in sentencing, identical to specification (4) in Table 2. All regressions contain controls for offense type, and dummies for each offense level and criminal history combination. Regressions also contain district by sentencing year, and sentencing month fixed effects, and standard errors are clustered at the district level. Race trends are included. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

While racial disparities in sentence length have increased as a whole, a more disaggregated analysis reveals that the growing racial disparities after *Booker* do not appear uniformly across all offenses. Table 2.9 presents results on sentence lengths disaggregated into the most prevalent seven offenses, which comprise 84% of all offenses in the dataset. Racial disparities increase significantly among defendants convicted of drug trafficking offenses, controlling for primary type of drug, and fraud offenses. Black and Hispanic defendants convicted of these offenses receive 1.5-2 months longer in prison compared to their white counterparts in the aftermath of *Booker*.

TABLE 2.9. SENTENCE LENGTHS BY MAJOR OFFENSE CATEGORIES

	(1) Drugs	(2) Immigration	(3) Firearms	(4) Fraud	(5) Bank Robbery	(6) Larceny	(7) Forgery
Booker*Black	2.354*** (0.804)	0.0632 (0.473)	0.475 (0.853)	1.487*** (0.408)	-2.248 (1.919)	0.379 (0.791)	0.278 (0.532)
Booker*Hispanic	1.275* (0.674)	0.643 (0.503)	1.167 (1.434)	1.996*** (0.603)	-7.224** (3.042)	-0.0837 (1.258)	-0.331 (0.561)
Booker*Other	0.177 (1.415)	-0.112 (1.007)	4.832* (2.552)	0.703 (1.018)	-3.872 (3.737)	1.627* (0.903)	1.623 (1.885)
Black	4.265*** (0.647)	-0.00646 (0.468)	1.791*** (0.613)	0.203 (0.172)	0.908 (0.968)	-0.250 (0.230)	-0.461 (0.332)
Hispanic	3.594*** (0.454)	-0.173 (0.484)	-0.926 (0.912)	-0.672*** (0.224)	2.745 (2.477)	-0.474 (0.450)	-0.471 (0.317)
Other	1.686 (1.447)	1.047 (1.116)	0.319 (2.175)	-0.137 (0.337)	-1.195 (2.698)	-0.167 (0.445)	0.183 (0.647)
Booker	-3.927** (1.716)	-0.859* (0.506)	-2.090 (2.520)	-2.721** (1.159)	-9.703* (5.604)	0.781 (1.380)	0.859 (2.751)
Observations	299,687	123,882	69,241	59,130	21,704	12,222	9,546
R-squared	0.752	0.812	0.720	0.749	0.687	0.795	0.785

Notes: Data is from the USSC from 1994-2009. Column 1 includes controls for primary drug type. All regressions contain dummies for each offense level and criminal history combination. Regressions also contain district by sentencing year, and sentencing month fixed effects, and standard errors are clustered at the district level. Race specific trends are excluded because of limited variation, but magnitudes are unchanged when race trends are included. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

2.6.2. Departures from the Guidelines

Table 2.10 presents results on how *Booker* impacted departures from the Guidelines. Column 1 replicates specification 4 of the sentence length results from Table 2.6. Column 2 indicates that post *Booker*, black defendants are sentenced at greater rates above range than white defendants, approximately 2%. However, conditional on above range sentencing, black defendants receive about the same number of months above range compared to white defendants.

TABLE 2.10. SENTENCING DEPARTURES FROM THE GUIDELINES

	(1) Sentence	(2) Above Range	(3) Months Above	(4) Below Range	(5) Months Below	(6) Within Range	(7) Months Within
Booker*Black	1.639** (0.690)	0.0186*** (0.00358)	-0.395 (3.255)	-0.0161** (0.00743)	-10.74 (6.663)	-0.00247 (0.00624)	0.879*** (0.212)
Booker*Hispanic	1.098** (0.539)	0.00694* (0.00414)	4.562 (3.083)	0.0110 (0.0111)	0.854 (4.852)	-0.0179* (0.00949)	0.260* (0.145)
Booker*Other	-0.168 (1.235)	-0.00503 (0.00592)	-0.274 (7.719)	0.0226 (0.0148)	-3.812 (9.318)	-0.0176 (0.0153)	-0.149 (0.392)
Black	3.185*** (0.591)	0.00132 (0.00312)	1.981 (2.508)	-0.0541*** (0.00709)	4.134 (4.781)	0.0527*** (0.00723)	-0.358*** (0.116)
Hispanic	1.308** (0.524)	-0.00998*** (0.00270)	-5.831** (2.226)	-0.0658*** (0.00952)	-0.311 (3.917)	0.0757*** (0.00947)	0.0262 (0.124)
Other	3.177*** (1.055)	0.0151*** (0.00471)	4.535 (5.593)	-0.0667*** (0.0242)	1.999 (6.727)	0.0516** (0.0240)	0.291 (0.232)
Booker	-2.593** (1.129)	0.00473 (0.00764)	1.460 (4.447)	0.0892*** (0.0141)	13.58** (5.247)	-0.0940*** (0.0140)	-0.230 (0.280)
Race Trends?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	679,159	679,159	41,478	679,159	255,776	679,159	381,901
R-squared	0.741	0.168	0.239	0.193	0.727	0.164	0.981

Notes: Data is from the USSC from 1994-2009. All regressions contain controls for offense type, and dummies for each offense level and criminal history combination. Regressions also contain district by sentencing year, and sentencing month fixed effects, and standard errors are clustered at the district level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Column 4 shows that below range departures increase generally post *Booker* by over 8% for all defendants. The high rate of below range departures following *Booker* may be the result of judicial discontent with the mandatory Guidelines regime. In a USSC survey of federal district judges in 2002, 30-40% of respondents stated that they believed that the Guidelines avoided unwarranted sentencing disparity only "Sometimes" or "Rarely." In a 2010 USSC survey of federal district judges after *Booker*, 65% of respondents indicated that they thought the departure policy statements in the Guidelines Manual were too restrictive.

While below range departures increase for all defendants in the aftermath of *Booker*, black offenders are significantly less likely to be sentenced below range compared to white defendants. Post *Booker*, black defendants are 1.6% less likely to be sentenced below range compared to similar white defendants. These results on below range departures are robust to excluding cases with statutory minimums.¹⁹

Finally, the last two columns indicate that rates of within range sentencing generally decreased by over 9% post *Booker*, but not differentially for black and white offenders. However, conditional on being sentenced within range, black offenders receive a 0.9 month longer sentence compared to their white counterparts post *Booker*. Recall that prior to *Booker*, judges were generally not allowed to consider factors such as defendant age, education, physical or mental problems, family, etc. in making sentencing decisions, except for within range sentences. The finding that disparities increase after *Booker* even for the subset of within range sentences suggests that disparities are not solely driven by the ability of judges to consider various unobservable factors in the aftermath of *Booker*.

Hispanic defendants face similar increases in disparities in departures from the Guidelines compared to similar white defendants. After *Booker*, Hispanic defendants are about 0.7% more likely to be sentenced above range, 1.8%

¹⁹Although not presented here, the differential rates of below range departures are not driven by government sponsored departures, but attributable to judicial departures.

less likely to be sentenced within range, and conditional on being sentenced within range, receive a 0.3 month longer sentence compared to white defendants. Thus, it appears that the increased racial disparities in sentencing between defendants occurs in the differential application of upward and downward departures, as well as disparate sentence lengths for within range sentences.²⁰

TABLE 2.11. OTHER SENTENCING OUTCOMES

	(1) Incarceration	(2) Probation Length	(3) Supervised Release Receipt	(4) Supervised Release
Booker*Black	0.00525 (0.00392)	-0.146 (0.635)	0.00154 (0.00132)	-1.522*** (0.316)
Booker*Hispanic	0.00806* (0.00437)	2.119* (1.109)	-0.00129 (0.00283)	-1.433*** (0.322)
Booker*Other	0.00260 (0.00870)	-1.894* (0.998)	-0.00723* (0.00376)	-2.529*** (0.553)
Black	0.0172*** (0.00319)	-0.269 (0.512)	-0.00249** (0.00123)	1.771*** (0.224)
Hispanic	0.00776* (0.00447)	-5.952*** (0.863)	-0.000426 (0.00248)	0.124 (0.246)
Other	-0.00533 (0.00865)	1.265 (0.963)	0.0145** (0.00650)	1.085** (0.441)
Booker	-0.0212*** (0.00733)	-2.400*** (0.855)	0.00861* (0.00499)	0.932* (0.501)
Race Trends?	Yes	Yes	Yes	Yes
Observations	817,222	137,499	678,699	666,846
R-squared	0.468	0.356	0.139	0.454

Notes: Data is from the USSC from 1994-2009. All regressions contain controls for offense type, and dummies for each offense level and criminal history combination. Regressions also contain district by sentencing year, and sentencing month fixed effects, and standard errors are clustered at the district level. Race trends are included. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

2.6.3. Robustness Checks for Increasing Racial Disparities

The previous results identify growing racial disparities in sentence length and departures from the Guidelines after *Booker*. One may be concerned that the increase in racial disparities after *Booker* is driven by harsher treatment of other characteristics that are associated with black defendants. For instance, if black defendants disproportionately have lower educational attainment, and judges take a harsher sentencing stance on less educated defendants post *Booker*, racial disparities may mechanically increase. To account for possible disparities driven by other demographic and crime characteristics, I include full interactions between the *Booker* indicator and a variety of relevant observables. In column 1 of Table 2.12, I replicate column 1 from Table 2.6 to show the baseline results. In column 2, I account for potential disparities post *Booker* based on defendant citizenship status, educational attainment, number of dependents, gender and age. In column 3, I account for possible disparities attributable to final offense level and criminal history

²⁰ An analysis of other sentence outcomes is presented in Table 2.11. Black offenders are generally more likely to be incarcerated compared to white offenders, but the differential in incarceration rates does not change post *Booker*. Probation lengths by defendant race do not change significantly post *Booker*. However, length of supervised release (served after imprisonment), changes substantially. Black defendants generally receive almost 2 months longer of supervised release, compared to similar white defendants. Post *Booker*, black and Hispanic defendants receive about 1.5 months less of supervised release compared to white defendants. The divergent changes in racial disparities in sentence length and supervised release length after *Booker* may be a result of judges replacing actual sentences for supervised release time for black and Hispanic defendants.

category. Finally, column 4 also accounts for disparities attributable to offense type.

TABLE 2.12. DISPARITIES IN SENTENCE LENGTH BY OTHER CHARACTERISTICS

	(1)	(2)	(3)	(4)
	Sentence	Sentence	Sentence	Sentence
Booker*Black	2.373*** (0.595)	2.024*** (0.562)	1.680*** (0.537)	1.326*** (0.499)
Booker*Hispanic	1.687*** (0.446)	1.561*** (0.452)	1.226*** (0.427)	0.975** (0.430)
Booker*Other	2.711*** (0.986)	2.699*** (0.987)	2.840*** (0.964)	2.431** (1.067)
Booker*Non US Citizen		-0.650* (0.334)	-0.189 (0.310)	-0.341 (0.362)
Booker*HS Grad		-0.461* (0.272)	-0.413 (0.263)	-0.254 (0.261)
Booker*Some College		-0.940*** (0.357)	-0.699** (0.331)	-0.303 (0.312)
Booker*College Grad		-2.544*** (0.548)	-1.902*** (0.491)	-1.639*** (0.489)
Booker*# Dependents		-0.0530 (0.0594)	-0.0983 (0.0612)	-0.133** (0.0613)
Booker*Female		-0.734** (0.362)	-0.125 (0.323)	0.159 (0.311)
Booker*Age		-0.0151 (0.0113)	-0.00991 (0.0105)	-0.00792 (0.0110)
Booker*Criminal History 2			1.816*** (0.343)	1.444*** (0.369)
Booker*Criminal History 3			2.218*** (0.401)	1.684*** (0.431)
Booker*Criminal History 4			1.664*** (0.458)	0.855* (0.470)
Booker*Criminal History 5			2.528*** (0.611)	1.528*** (0.581)
Booker*Criminal History 6			-0.447 (0.751)	-1.092 (0.735)
Observations	679,159	679,159	679,159	679,159
R-squared	0.741	0.741	0.742	0.742

Notes: Data is from the USSC from 1994-2009. All regressions contain controls for offense type, and dummies for each offense level and criminal history combination. Column 1 replicates column 1 from Table 2 to show the baseline results. Column 2 includes interactions between defendant race and citizenship status, educational attainment, number of dependents, gender and age. Column 3 adds interactions between defendant race and final offense level and criminal history category. Finally, column 4 adds interactions between race and offense type. Regressions also contain district by sentencing year, and sentencing month fixed effects, and standard errors are clustered at the district level. Race specific trends are excluded because of limited variation, but magnitudes are unchanged when race trends are included. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Note that the significance of the coefficients on *Booker* interacted with defendant race remained unchanged in all 4 columns and but falls in magnitude. Racial differences in sentencing are not the only disparities that emerge after *Booker*. The results from Table 4 reveal growing disparities among defendants of different educational attainments. After *Booker*, defendants with some college and those with a college degree are sentenced to almost 2 months less, compared to their less educated counterparts. Furthermore, defendants with additional dependents face a slightly lower prison sentence compared to defendants with fewer dependents post *Booker*. In contrast, disparities do not increase by gender, age or citizenship status.

Fully accounting for disparities due to defendant offense level and criminal history category reveals additional disparities post *Booker*. The coefficients on offense level interacted with the *Booker* indicator are omitted because none

are statistically significant at the 10% level, suggesting that judges do not differentially sentence defendants with different offense severity post *Booker*. However, judges sentence defendants with higher levels of prior criminal activity more harshly post *Booker*. After *Booker*, defendants in criminal history categories 2, 3, 4, and 5 face an approximately 1.5 month longer sentence, compared to first time offenders in criminal history category 1. When column 4 includes additional interactions with offense type, none of the coefficients are significant and are thus excluded, suggesting that judges are not sentencing differentially across offenses in the aftermath of *Booker*.

To further test the robustness of the results, I explore whether increasing racial disparities may be mechanically driven by black defendants being less likely to show remorse for their crimes. I capture this through the court's decision to reduce a defendant's offense level by either two or three points through the acceptance of responsibility provision. I find that lack of remorse as proxied by acceptance of responsibility cannot explain the growing racial disparities in the aftermath of *Booker* (See Table 2.13). Overall, these results suggest that racial disparities are robust to differential treatment of defendants by other characteristics in the aftermath of *Booker*. Despite increasing disparities by educational attainment, family structure, and criminal history, racial disparities persist.

TABLE 2.13. ACCEPTANCE OF RESPONSIBILITY REDUCTION

	(1) 2 Point Reduction	(2) 3 Point Reduction	(3) Any Reduction
Booker*Black	0.00556 (0.00550)	-0.00299 (0.00475)	-0.00445 (0.00368)
Booker*Hispanic	0.00433 (0.00553)	0.000752 (0.00433)	-0.00140 (0.00391)
Booker*Other	0.0200** (0.00908)	0.0178 (0.0111)	0.0130* (0.00658)
Black	-0.0114*** (0.00397)	-0.0314*** (0.00349)	-0.0177*** (0.00270)
Hispanic	0.0174*** (0.00453)	-0.00817* (0.00459)	0.00215 (0.00416)
Other	-0.0206*** (0.00650)	-0.000458 (0.0104)	-0.00305 (0.00757)
Booker	0.000354 (0.0111)	-0.00246 (0.00775)	-0.000984 (0.00616)
Race Trends?	Yes	Yes	Yes
Observations	326,524	569,481	822,002
R-squared	0.596	0.399	0.272

Notes: Data is from the USSC from 1994-2009. All regressions contain controls for offense type, and dummies for each offense level and criminal history combination. Regressions also contain district by sentencing year, and sentencing month fixed effects, and standard errors are clustered at the district level. Race trends are included. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

2.6.4. How Constraining is Appellate Review? Evidence from *Rita*, *Gall*, and *Kimbrough*

Booker changed the legal landscape by invalidating the mandatory nature of the Guidelines, but the series of Supreme Court decisions that followed also changed the standard of appellate review. In the first two and half years after *Booker*, judges were no longer bound to the Guidelines, but still faced a high level of appellate scrutiny. Beginning in late 2007, the *Rita* presumption of reasonableness for within range sentences provided judges with a safe harbor

from appellate scrutiny. *Gall* and *Kimbrough* removed the presumption of unreasonableness for sentences outside the Guidelines range, further reducing the probability of reversal.

These differential changes in the increase in judicial discretion yield insights into the mechanisms to which judges respond. If judges are greatly bound by the rule-based nature of the Guidelines, one would expect to see large increases in disparities immediately after *Booker*. If judges are constrained by appellate review, the advisory nature of the Guidelines coupled with strict standards of review may still restrict judicial sentencing. Instead, judges constrained by appellate review would be most free to deviate in the aftermath of *Rita*, *Gall*, and *Kimbrough*.

To capture the dynamics in the aftermath of *Booker*, I replicate specification (1) using leads and lags in six month intervals for the five years prior and post *Booker*. These leads and lags are then interacted with defendant race to capture the change in disparities in that specific time period compared to the base period (1994-1999).

Figure 2.2 presents the results from a dynamic differences-in-differences specification where the dependent variable is sentence length in months.²¹ Figure 2.2 graphs the coefficients for the leads and lags interacted with a black race dummy, along with corresponding 95% confidence intervals, and shows a clear increasing sentencing gap between black and white defendants. The lack of a significant gap between black and white defendants in the five years prior to *Booker* suggests that preexisting trends cannot explain growing racial disparities.

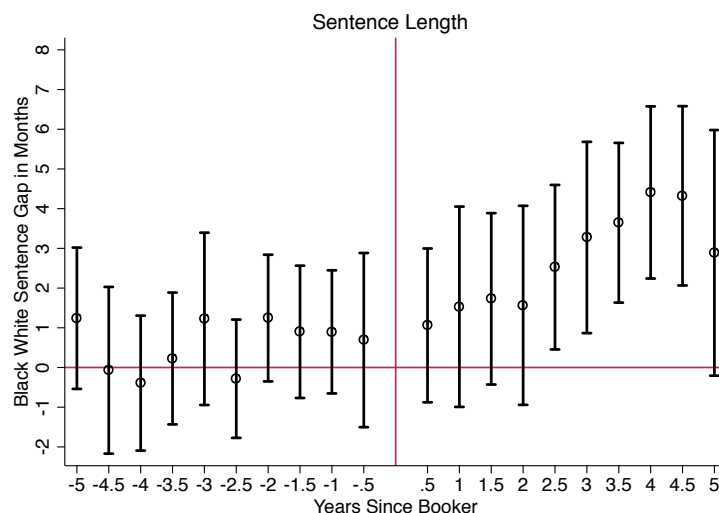


FIGURE 2.2. DYNAMICS OF BLACK WHITE GAP, SENTENCE LENGTH IN MONTHS

Notes: Data is from the USSC from 1994-2009. This figure shows coefficients from a dynamic DD regression identical to specification (1) in Table 2.6, but with leads and lags for the five years before and five years after *Booker*, interacted with defendant race. The coefficients represent the differential sentence lengths between black and white defendants, compared to the pre-period (1994-1999). Error bars represent 95% confidence intervals. Race specific trends are excluded because of limited variation, but magnitudes are unchanged when race trends are included.

Starting about two and a half years after *Booker*, black defendants appear to face a 2.5 month longer sentence compared to their white counterparts, and the sentencing disparity continues to rise over time. By four years after

²¹ Although not presented here, results for Hispanic defendants compared to white defendants show a similar, but less pronounced trend.

Booker, the sentencing gap increases to 4.4 months, almost a 10% increase in the average sentence length. (See Table 2.14 for results in table format). The fact that racial disparities are not significant in the immediate aftermath of *Booker* suggests that *de novo* review may have still been a binding constraint on judicial sentencing, even though the Guidelines were rendered advisory. The appearance of rising racial disparities approximately two and a half years after *Booker* coincide with *Rita*, *Gall* and *Kimbrough*, indicating that more deferential appellate review greatly affects judicial sentencing behavior.

TABLE 2.14. DYNAMIC SPECIFICATION, BLACK WHITE GAP

	(1) Black White Gap Sentence	(2) Black White Gap Above Range	(3) Black White Gap Below Range
55-60 Months Before	1.240 (0.908)	-0.000592 (0.00497)	-0.00776 (0.0111)
49-54 Months Before	-0.0703 (1.071)	-0.00332 (0.00549)	-0.00183 (0.0103)
43-48 Months Before	-0.393 (0.867)	-0.0109** (0.00508)	-0.00248 (0.00967)
37-42 Months Before	0.227 (0.847)	-0.00367 (0.00519)	-0.0115 (0.00797)
31-36 Months Before	1.225 (1.107)	-0.00616 (0.00469)	-0.00888 (0.00978)
25-30 Months Before	-0.283 (0.760)	-0.00755 (0.00515)	0.00738 (0.0115)
19-24 Months Before	1.245 (0.814)	-0.00599 (0.00468)	-0.00305 (0.0101)
13-18 Months Before	0.898 (0.850)	-0.0103** (0.00415)	-0.00122 (0.00928)
7-12 Months Before	0.897 (0.791)	-0.00564 (0.00451)	0.00753 (0.0105)
1-6 Months Before	0.690 (1.119)	-0.00257 (0.00568)	0.0103 (0.00923)
1-6 Months After	1.060 (0.988)	-0.00973* (0.00579)	-0.00192 (0.0122)
7-12 Months After	1.530 (1.287)	-0.00363 (0.00713)	0.00142 (0.0114)
13-18 Months After	1.729 (1.101)	-0.00556 (0.00535)	-6.17e-05 (0.0105)
19-24 Months After	1.566 (1.279)	-0.000684 (0.00636)	-0.0161 (0.0115)
25-30 Months After	2.526** (1.057)	0.00963 (0.00608)	-0.00910 (0.0109)
31-36 Months After	3.274*** (1.229)	0.0140** (0.00596)	-0.0204* (0.0104)
37-42 Months After	3.645*** (1.026)	0.0207*** (0.00499)	-0.0333*** (0.0123)
43-48 Months After	4.408*** (1.106)	0.0288*** (0.00639)	-0.0516*** (0.0112)
49-54 Months After	4.323*** (1.152)	0.0259*** (0.00702)	-0.0375*** (0.0124)
55-58 Months After	2.887* (1.578)	0.0348*** (0.00913)	-0.0223 (0.0147)
Observations	692,039	692,039	692,039
R-squared	0.741	0.169	0.193

Notes: Data is from the USSC from 1994-2009. Coefficients are for the differential outcome for black vs. white defendants from a dynamic DD regression identical to specification (1) in Table 2, but with leads and lags for the five years before and five years after *Booker*, interacted with defendant race. All regressions contain controls for offense type, and dummies for each offense level and criminal history combination. Regressions also contain district by sentencing year, and sentencing month fixed effects, and standard errors are clustered at the district level. Race trends are excluded. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Figure 2.3 captures the pattern in departures from the Guidelines, where the dependent variable is an indicator variable for an above range departure or below range departure. The gap in above range sentencing for black and white defendants appears starting around two and a half years after *Booker*, persists and grows larger. By five years after the *Booker* decision, black defendants are over 3.5% more likely to be sentenced above range compared to their white counterparts. Similarly, the gap in below range sentencing starts around three years after *Booker* and persists throughout the rest of the period, with black defendants over 5% less likely to be sentenced below range compared to

white defendants four years after *Booker*. Again, racial disparities in the rate of departures became more pronounced after *Rita*, *Gall* and *Kimbrough*, suggesting that judges are particularly responsive to standards of appellate review. I present evidence in the next section suggesting that the growing racial disparities are also attributable to the increasing number of judges appointed post *Booker*.

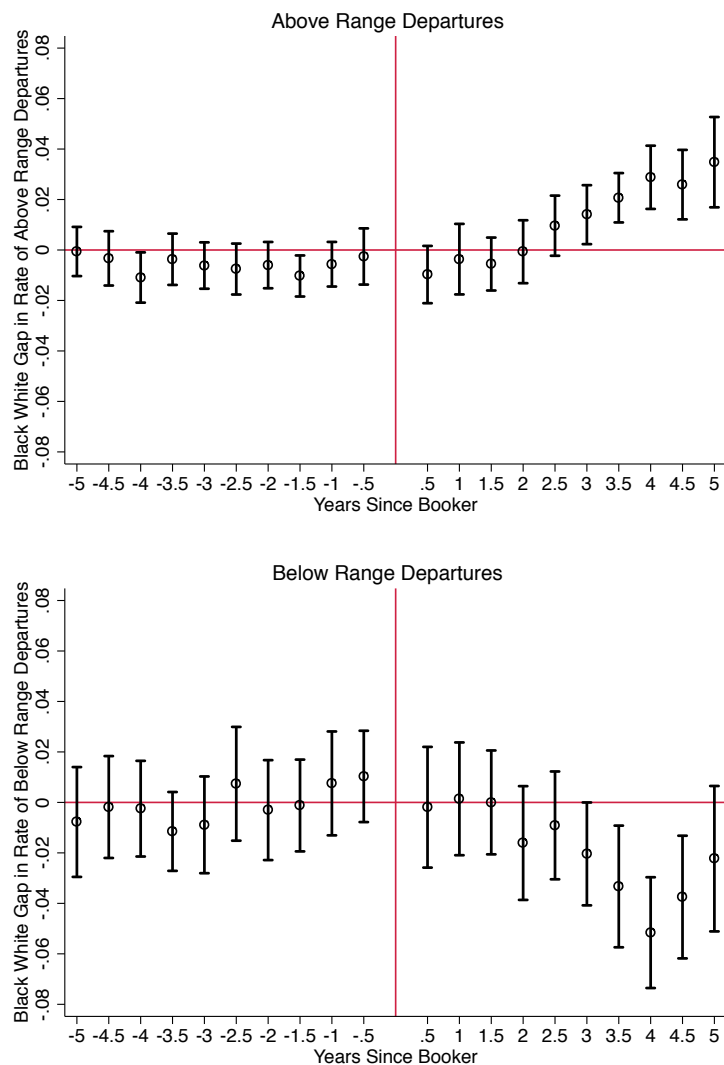


FIGURE 2.3. DYNAMICS OF BLACK WHITE DEPARTURE RATES

Notes: Data is from the USSC from 1994-2009. This figure shows coefficients from a dynamic DD regression identical to specification (1) in Table 2.6, but with leads and lags for the five years before and five years after *Booker*, interacted with defendant race. The coefficients represent the differential sentence lengths between black and white defendants, compared to the pre-period (1994-1999). Error bars represent 95% confidence intervals. Race specific trends are excluded because of limited variation, but magnitudes are unchanged when race trends are included.

2.6.5. Free at Last? Effects of Judicial Sentencing Philosophies and Experience

While disparities in sentencing outcomes increased in the wake of *Booker*, the response to increased judicial discretion may differ by judge sentencing philosophies and experience. In particular, judges appointed before *Booker*

may sentence differently compared to judges appointed after *Booker*. Judges with substantial experience sentencing under the mandatory Guidelines regime may become acculturated to the Guidelines, and less likely to change their sentencing practices in the aftermath of *Booker*.

Since *Booker*, there have been 190 confirmed judicial appointments to US district courts, 93 new judges up to the end of the fiscal sentencing year 2009.²² The judges appointed prior to 2009 were appointed by President George W. Bush, and the remaining judges by President Barack Obama. However, all Obama appointees began active service following the end of the fiscal year 2009, so this paper cannot identify the sentencing patterns of new Democratic appointed judges. Within the matched data from 2000-2009, post *Booker* appointed judges have sentenced a growing share of criminal defendants, to almost 10% of cases in fiscal year 2009.

Recall that random assignment of cases to judges is necessary in order to compare sentencing practices of judges within a district court. According to the Administrative Office of the United States Courts, “[t]he majority of courts use some variation of a random drawing” as prescribed by local court orders. However, random assignment may be violated in some instances. For example, senior status judges with reduced caseloads may select the type of cases they hear during the year, and some courts assign certain types of cases to particular judges.

To exclude senior status judges who may not obtain cases through a random assignment process, I drop judges who were formally retired prior to 2000, and judges and district courthouses with annual caseloads of less than 25 cases. To ensure that I only include courthouses with random assignment of cases, I then test for random assignment by district courthouse using the matched USSC, TRAC, and Federal Judicial Center data from 2000-2009, for a set of five predetermined defendant characteristics: gender, age, a black race indicator, number of dependents, and an indicator for less than a high school degree. For each of the five defendant characteristics, I regress the characteristic on district courthouse by sentencing year fixed effects, sentencing month fixed effects and judge fixed effects. I test the hypothesis of no judge effects (the null hypothesis) using an F-test for whether the judge fixed effects are equal to zero using seemingly unrelated regression (SUR) following Autor and Houseman (2010). P-values for these tests by district courthouse are presented in Table 2.15. I drop all courthouses with F-test p-values less than 0.05, but results are robust to other cutoffs. The subsample of district courts with random case assignment includes 72 courts representing about 50% of the cases from 2000-2009.²³

²²Nine judges were commissioned in 2005, 26 commissioned in 2006, 32 in 2007, and 26 in 2008. Post *Booker* appointed judges are now active in 53 district courts, some comprising up to 75% of the active bench within a court.

²³Table 2.16 presents the results of the core specification from Table 2 using the random sample and full matched sample.

TABLE 2.15. RANDOMIZATION TESTS 2000-2009

District Court	No. Obs.	p-value
ME (0)	1,668	0.1438
MA (1)	4,042	0.1054
NH (2)	1,617	0.9844
PR (4)	6520	0.2674
CT** (5)	664	0.0000
NY North - Syracuse (6)	1,148	0.1074
NY East** (7)	12,447	0.0004
NY South - White Plains (8)	1,338	0.4336
NY West - Rochester (9)	1,166	0.6226
VT (10)	1,400	0.2379
DE (11)	641	0.3831
NJ -Trenton (12)	476	0.2983
PA East** (13)	6,411	0.0000
PA Middle - Scranton (14)	969	0.6837
PA Middle - Williamsport (14)	234	0.2071
PA West - Erie (15)	609	0.0521
PA West - Pittsburgh (15)	2,917	0.0645
MD (16)	5,569	0.0631
NC East - Southern (17)	608	0.3847
NC Middle (18)	3,205	0.08086
NC West** (19)	5,563	0.0000
SC** (20)	8,848	0.0000
VA East -Alexandria (22)	4,500	0.3178
VA East -Norfolk (22)	1,105	0.1658
VA East -Newport News (22)	743	0.0662
VA West (23)	3,123	0.3250
WV North - Martinsburg (24)	639	0.4091
WV South (25)	1,778	0.0932
AL North** (26)	1,430	0.0189
AL Middle (27)	904	0.3242
AL South (28)	3,132	0.0702
FL North (29)	2,718	0.5783
FL Middle - Ft. Myers (30)	923	0.3824
FL Middle - Ocala (30)	465	0.3128
FL South - Ft. Pierce (31)	3,299	0.0541
FL South - Ft. Lauderdale (31)	649	0.2485
GA North** (32)	5,823	0.0000
GA Middle (33)	2,064	0.1396
LA East (35)	3,117	0.0606
LA West (36)	1,686	.6360
MS North (37)	925	0.4247
MS South (38)	3,057	0.0564
TX North - Forth Worth (39)	2,027	0.2386
TX East (40)	6,563	0.5598
TX South - Brownsville (41)	10,112	0.3364
TX South - Corpus Christi (41)	6,679	0.2767
TX South - Laredo (41)	19,079	0.6244
TX South - McAllen (41)	12,739	0.1093

TABLE 2.15. RANDOMIZATION TESTS 2000-2009 (CONTINUED)

TX West - Del Rio (42)	7,098	0.3500
TX West - Midland-Odessa (42)	3,567	0.4120
KY East - Covington (43)	717	0.5872
KY East - Pikeville (43)	139	0.0966
KY East - Lexington (43)	1,993	0.8694
KY West (44)	1,746	0.1114
MI East - Bay City (45)	458	0.4009
MI East - Flint (45)	673	0.3014
MI West (46)	3,313	0.0961
OH North - Toledo (47)	1,014	0.2105
OH South - Dayton (48)	1,300	0.9115
TN East (49)	5,200	0.0705
TN Middle** (50)	1,938	0.0126
TN West - Eastern (51)	831	0.3998
IL North - Rockford (52)	624	0.8929
IL Central (53)	2,618	0.1283
IL South (54)	2,736	0.1296
IN North - South Bend (55)	954	0.2764
IN North - Fort Wayne (55)	530	0.0741
IN South (56)	2,004	0.3266
WI East - Milwaukee (57)	2,206	0.4223
WI West (58)	1,571	0.1123
AR East (60)	2,739	0.1631
AR West** (61)	1,098	0.0001
IA North (62)	2,413	0.0561
IA South (63)	2,684	0.8265
MN** (64)	4,815	0.0001
MO East (65)	8,203	0.0785
MO West (66)	6,764	0.1191
NE - Omaha (67)	2,323	0.0532
ND (68)	1,888	0.2250
SD - Aberdeen (69)	309	0.1479
SD - Pierre (69)	1,010	0.8757
AZ - Tuscon (70)	23,677	0.0961
AZ - Yuma (70)	2,449	0.3392
CA North (71)	3,045	0.1970
CA East (72)	8,094	0.0646
CA Central - Riverside (73)	157	0.4520
CA South - El Centro (74)	8,664	0.3442
CA South - Yuma (74)	89	0.3502
HI** (75)	3,351	0.0012
ID (76)	1,526	0.0544
MT - Missoula (77)	516	0.1698
MT - Great Falls (77)	1,003	0.2206
NV (78)	4,867	0.6549
OR - Eugene (79)	954	0.2261
OR - Medford (79)	434	0.6618
WA East - Spokane (80)	1,401	0.3100
WA West** (81)	5,302	0.0001

TABLE 2.15. RANDOMIZATION TESTS 2000-2009 (CONTINUED)

CO** (82)	4,582	0.0000
KS (83)	5,509	0.2031
NM (84)	24,019	0.2924
OK North (85)	1,279	0.3240
OK East (86)	736	0.9312
OK West** (87)	1,809	0.0001
UT (88)	5,276	0.9421
WY** (89)	1,565	0.0002
DC (90)	346	0.5720
AK (95)	1,218	0.1105
LA Middle** (96)	1,112	0.0263

Notes: Data is from the matched USSC, TRAC, Federal Judicial Center data from 2000-2009. I drop judges who retired or were terminated prior to 2000, and judges and district offices with an annual caseload of less than 25. For each district court, I control for district office by sentencing year, sentencing month, and judge fixed effects. P-values reported test whether judge fixed effects differ significantly from zero and are from a seemingly unrelated regression (SUR) on five defendant characteristics: defendant gender, age, black race indicator, number of dependents, and less than high school indicator. ** indicates dropped courthouses.

TABLE 2.16. MAIN RESULTS USING JUDGE MATCHED DATA

	(1) Random Sample Sentence	(2) Full Matched Sample Sentence
Booker*Black	1.994*** (0.691)	2.486*** (0.546)
Booker*Hispanic	0.732 (0.563)	1.076** (0.460)
Booker*Other	0.170 (0.880)	1.535 (1.035)
Black	2.924*** (0.584)	2.315*** (0.404)
Hispanic	2.155*** (0.598)	1.270*** (0.467)
Other	3.070*** (0.713)	1.932* (1.040)
Booker	-2.897*** (1.073)	-2.811*** (1.045)
Non US Citizen	0.433 (0.496)	1.283** (0.502)
HS Grad	-0.701*** (0.215)	-0.555*** (0.186)
Some College	-1.434*** (0.310)	-1.762*** (0.185)
College Grad	-2.119*** (0.378)	-2.068*** (0.247)
# Dependents	-0.250*** (0.0713)	-0.236*** (0.0477)
Female	-4.996*** (0.515)	-5.035*** (0.508)
Age	0.241*** (0.0606)	0.190*** (0.0446)
Age ²	-0.00263*** (0.000704)	-0.00208*** (0.000503)
Mandatory Min	22.13*** (2.086)	21.41*** (1.841)
Observations	214,136	478,834
R-squared	0.784	0.754

Notes: Data is from the matched USSC, TRAC, Federal Judicial Center data from 2000-2009. Column 1 replicates column 1 from Table 2 using the sample of random courts. Column 2 replicates column 1 of Table 2 using the full matched sample. All regressions contain controls for offense type, and dummies for each offense level and criminal history combination. Regressions also contain district by sentencing year, and sentencing month fixed effects, and standard errors are clustered at the district level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table 2.17 presents the results, using this subsample of district courts, of specification (2) with an interaction between defendant race, the *Booker* indicator, and an indicator variable equal to one for judges appointed post *Booker*, in addition to the interaction between defendant race and the *Booker* indicator.²⁴ The triple interaction terms measures the different sentencing practices of post *Booker* appointed judges on disparities in sentencing, compared to pre *Booker* judges in the aftermath of *Booker*. Column 1 presents results for sentence length. The coefficients of the *Booker* indicator interacted with defendant race indicate that racial disparities increase by 1.7 months between black and white defendants after *Booker*, but particularly for post *Booker* appointed judges. These “new” judges sentence black defendants to an additional 5.4 months in prison compared to similar white defendants, relative to their colleagues.

²⁴Note that because all “new” judges were appointed after *Booker*, in this instance, the triple interaction is identical to an interaction between defendant race and “new” judge.

TABLE 2.17. SENTENCING PATTERNS FOR POST *Booker* JUDGES
SUBSAMPLE OF RANDOM DISTRICTS

	(1) Sentence	(2) Above Range	(3) Months Above	(4) Below Range	(5) Months Below	(6) Within Range	(7) Months Within
Post Booker Judge	-1.145 (1.727)	-0.00554 (0.0113)	-0.194 (7.906)	-0.00260 (0.0166)	0.987 (1.355)	0.00814 (0.0142)	-0.759** (0.293)
Post Booker*Black	5.440** (2.587)	0.0200 (0.0149)	6.293 (6.918)	-0.0101 (0.0246)	-2.103 (1.751)	-0.00994 (0.0201)	1.243** (0.501)
Post Booker*Hispanic	-0.625 (1.574)	0.0146 (0.0228)	-7.618 (7.826)	0.00725 (0.0289)	-0.797 (1.826)	-0.0219 (0.0311)	0.838** (0.365)
Post Booker*Other	0.218 (2.625)	0.0387 (0.0306)	-12.89 (12.47)	0.0343 (0.0642)	2.548 (2.407)	-0.0730 (0.0607)	2.862 (2.300)
Booker	-2.835** (1.074)	-0.00253 (0.0108)	3.104 (5.495)	0.0849*** (0.0199)	-0.757 (1.137)	-0.0824*** (0.0211)	-0.303 (0.369)
Booker*Black	1.693** (0.696)	0.0145*** (0.00443)	4.988 (3.814)	-0.0144* (0.00734)	-1.061 (0.689)	-0.000126 (0.00715)	0.616*** (0.190)
Booker*Hispanic	0.744 (0.547)	0.000522 (0.00420)	-0.830 (2.757)	-0.0123 (0.0139)	0.0802 (0.466)	0.0118 (0.0129)	0.357*** (0.130)
Booker*Other	0.154 (0.874)	0.00403 (0.00914)	5.574 (6.971)	-0.0118 (0.0223)	0.415 (0.966)	0.00775 (0.0197)	-0.545 (0.358)
Observations	214,136	214,136	13,091	214,136	82,215	214,136	118,612
R-squared	0.784	0.194	0.367	0.244	0.736	0.202	0.985

Notes: Data is from the matched USSC, TRAC, Federal Judicial Center data from 2000-2009 for courts with random assignment, excluding judges who formally retired prior to 2000. All regressions contain controls for offense type, and dummies for each offense level and criminal history combination. Regressions also contain district office by sentencing year, district court fixed effects, sentencing month fixed effects, and standard errors are clustered at the district level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Column 2 indicates that all judges are associated with greater rates of above range departures for black defendants compared to white defendants. Column 4 also indicates different rates of below range departures for black and white defendants after *Booker* for all judges. As shown in column 7, conditional on within range sentencing, all judges sentence black defendants to about 0.6 months longer in prison and Hispanic defendants 0.3 months longer in prison compared to white defendants. However, the black-white sentence gap for within range sentences is 1.2 months larger for post *Booker* judges compared to pre *Booker* appointed judges. Similarly, the Hispanic-white sentence gap for within range sentences is 0.8 months larger for post *Booker* judges compared to their pre *Booker* colleagues.

These results indicate that post *Booker* appointed judges exhibit greater racial disparities in their sentencing patterns than their pre *Booker* colleagues, even within the same district courthouse.²⁵

²⁵Results are robust to using the full matched sample. See Table 2.18.

TABLE 2.18. SENTENCING PATTERNS FOR POST *Booker* JUDGES
FULL MATCHED SAMPLE

	(1) Sentence	(2) Above Range	(3) Months Above	(4) Below Range	(5) Months Below	(6) Within Range	(7) Months Within
Post Booker Judge	0.694 (1.097)	-0.00483 (0.00601)	-3.645 (4.444)	-0.0295* (0.0164)	-2.806 (4.064)	0.0343** (0.0162)	-0.0298 (0.258)
Post Booker Judge*Black	4.197** (1.776)	0.0234* (0.0135)	2.973 (4.455)	-0.0109 (0.0167)	-4.228 (6.711)	-0.0125 (0.0196)	1.004** (0.442)
Post Booker Judge*Hispanic	-0.147 (1.059)	0.0119 (0.0133)	-2.594 (4.246)	0.0138 (0.0226)	-1.786 (4.525)	-0.0257 (0.0241)	0.305 (0.290)
Post Booker Judge*Other	-4.157** (1.628)	-0.00702 (0.0139)	-9.901 (6.364)	0.0919*** (0.0318)	-7.199 (4.882)	-0.0849** (0.0343)	0.627 (0.845)
Booker	-2.763** (1.051)	0.00897 (0.00697)	-1.482 (6.409)	0.101*** (0.0149)	0.770 (4.228)	-0.110*** (0.0147)	-0.179 (0.276)
Booker*Black	2.296*** (0.542)	0.0141*** (0.00294)	4.203* (2.278)	-0.0170** (0.00647)	8.955** (3.707)	0.00296 (0.00566)	0.534*** (0.187)
Booker*Hispanic	1.071** (0.454)	0.000226 (0.00383)	2.678 (2.181)	-0.00414 (0.00935)	3.231 (2.743)	0.00391 (0.00834)	0.328*** (0.0911)
Booker*Other	1.717 (1.035)	0.00394 (0.00461)	0.351 (5.212)	-0.00673 (0.0108)	17.03* (10.24)	0.00279 (0.0106)	-0.000959 (0.233)
Observations	478,834	478,834	27,743	478,834	184,986	478,834	266,102
R-squared	0.754	0.169	0.274	0.204	0.842	0.172	0.983

Notes: Data is from the matched USSC, TRAC, Federal Judicial Center data from 2000-2009. All regressions contain controls for offense type, and dummies for each offense level and criminal history combination. Regressions also contain district by sentencing year, and sentencing month fixed effects, and standard errors are clustered at the district level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Given that cases are randomly assigned within a district, it is unlikely that these post *Booker* judges were assigned cases in which black defendants deserved longer sentences compared to their observably similar white counterparts. Moreover, greater racial disparities among post *Booker* appointed judges cannot be explained by the fact that these judges were appointed by George W. Bush. In Table 2.19, I include all interactions between defendant race, the *Booker* dummy variable, and an indicator variable for pre *Booker* Bush appointees. These controls allow me to compare the sentencing patterns of post *Booker* judges to their pre *Booker* appointed counterparts. The coefficient on Pre Booker Bush Judge indicates that pre *Booker* Bush appointees were generally 4% less likely to sentence below range for all defendants compared to their colleagues, but with no changes in sentencing practices in the aftermath of *Booker*.

TABLE 2.19. SENTENCING FOR POST *Booker* JUDGES - COMPARISON TO PRE *Booker* BUSH APPOINTEES
SUBSAMPLE OF RANDOM DISTRICTS

	(1) Sentence	(2) Above Range	(3) Months Above	(4) Below Range	(5) Months Above	(6) Within Range	(7) Months Within
Post Booker Judge	-1.002 (1.702)	-0.00562 (0.0115)	-0.424 (7.811)	-0.00969 (0.0163)	-1.785 (6.092)	0.0153 (0.0148)	-0.741** (0.297)
Post Booker Judge*Black	5.694** (2.621)	0.0198 (0.0150)	6.018 (6.666)	-0.0140 (0.0246)	0.886 (8.282)	-0.00584 (0.0194)	1.245** (0.496)
Post Booker Judge*Hispanic	-0.712 (1.568)	0.0147 (0.0230)	-7.757 (7.942)	0.0136 (0.0282)	-3.067 (5.560)	-0.0283 (0.0309)	0.812** (0.371)
Post Booker Judge*Other	0.503 (2.743)	0.0389 (0.0311)	-11.31 (13.03)	0.0307 (0.0633)	0.122 (6.237)	-0.0696 (0.0595)	2.684 (2.311)
Booker	-2.513** (1.087)	-0.00116 (0.0106)	0.534 (5.492)	0.0815*** (0.0193)	-1.134 (3.183)	-0.0803*** (0.0203)	-0.259 (0.383)
Booker*Black	1.658** (0.797)	0.0151*** (0.00518)	6.558* (3.860)	-0.0102 (0.00798)	6.725 (7.028)	-0.00491 (0.00757)	0.649*** (0.197)
Booker*Hispanic	0.766 (0.551)	0.000747 (0.00454)	-0.0243 (3.160)	-0.0118 (0.0147)	3.987 (3.042)	0.0111 (0.0136)	0.378*** (0.131)
Booker*Other	-0.231 (0.963)	0.00186 (0.00911)	5.001 (6.341)	-0.00783 (0.0202)	1.725 (4.334)	0.00597 (0.0186)	-0.272 (0.415)
Pre Booker Bush	1.219 (1.161)	0.00324 (0.00614)	-10.94* (6.186)	-0.0446** (0.0198)	-4.095 (2.861)	0.0414* (0.0219)	0.154 (0.242)
Pre Booker Bush*Black	2.315 (1.712)	0.00512 (0.00988)	9.985 (8.224)	-0.00139 (0.0176)	6.167 (8.236)	-0.00373 (0.0192)	0.318 (0.263)
Pre Booker Bush*Hispanic	0.0911 (1.398)	0.00169 (0.00617)	4.016 (6.225)	0.0361 (0.0238)	2.602 (2.934)	-0.0378 (0.0244)	-0.0322 (0.215)
Pre Booker Bush*Other	-0.567 (2.259)	-0.0100 (0.0104)	6.193 (18.15)	0.0121 (0.0353)	3.751 (5.223)	-0.00207 (0.0361)	0.481 (0.386)
Booker*Pre Booker Bush	-0.725 (1.241)	-0.00357 (0.00741)	9.820 (7.513)	0.0210 (0.0189)	5.753 (3.550)	-0.0174 (0.0228)	-0.109 (0.318)
Booker*Pre Booker Bush*Black	-1.375 (1.758)	-0.00575 (0.0124)	-10.94 (10.18)	-0.0141 (0.0193)	-6.872 (9.353)	0.0198 (0.0214)	-0.305 (0.453)
Booker*Pre Booker Bush*Hispanic	-0.124 (1.433)	-0.00171 (0.00765)	-5.195 (6.764)	-0.0226 (0.0200)	-4.572 (3.590)	0.0244 (0.0228)	-0.0529 (0.297)
Booker*Pre Booker Bush*Other	1.375 (1.626)	0.0108 (0.0135)	-0.555 (23.78)	-0.0206 (0.0386)	-2.776 (7.305)	0.00984 (0.0423)	-1.020 (0.646)
Observations	214,136	214,136	13,091	214,136	82,432	214,136	118,612
R-squared	0.784	0.194	0.368	0.244	0.919	0.202	0.985

Notes: Data is from the matched USSC, TRAC, Federal Judicial Center data from 2000-2009 for courts with random assignment, excluding judges who formally retired prior to 2000. All regressions contain controls for offense type, and dummies for each offense level and criminal history combination. Regressions also contain district by sentencing year, and sentencing month fixed effects, and standard errors are clustered at the district level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table 2.19 also indicates that the coefficients on Post Booker Judge and its interactions with defendant race remained unchanged from those presented in Table 5, confirming that the sentencing patterns of post *Booker* appointed judges are not attributable to the fact that these judges are George W. Bush appointees. Although not presented here, new judges in earlier cohorts also do not sentence differently from their more experienced colleagues, either before or after *Booker*, indicating that judge experience alone cannot explain inter-judge differences in sentencing. Instead, the results suggest that experience sentencing under a mandatory Guidelines regime may drive the differential sentencing patterns between pre and post *Booker* appointed judges.

Different sentencing philosophies and practices between judges may not only be driven by experience under a mandatory Guidelines regime, but other personal preferences. To proxy for sentencing philosophies, I replicate the regressions in Table 5 with additional controls for judge gender, race, political affiliation, and an indicator for whether the judge was appointed prior to the adoption of the Guidelines. Table 2.20 shows that post *Booker* appointed judges are still the main source of increasing racial disparities. However, other judicial demographic characteristics are also associated with certain sentencing patterns. Table 2.20 shows that female judges sentence all defendants to 2.4 months less in prison after *Booker* compared to their male colleagues. Table A15 also suggests that black judges were about 6% less likely to sentence defendants of other races below range prior to *Booker*, but reversed this practice in the

aftermath of *Booker*. Also striking are the different sentencing practices of Democratic and Republican appointed judges. Democratic judges are 2.1% more likely than Republican judges to depart downwards from the Guidelines, and even conditional on sentencing within range, issue a sentence to all defendants that is 0.4 months less compared to their Republican colleagues.

TABLE 2.20. SENTENCING PATTERNS BEFORE AND AFTER *Booker*, BY JUDICIAL DEMOGRAPHICS
SUBSAMPLE OF RANDOM DISTRICTS

	(1) Sentence	(2) Above Range	(3) Months Above	(4) Below Range	(5) Months Above	(6) Within Range	(7) Months Within
Post Booker Judge	-1.441 (1.759)	-0.00562 (0.0114)	-0.498 (8.488)	0.0136 (0.0174)	-2.160 (5.945)	-0.00794 (0.0148)	-0.754** (0.319)
Post Booker Judge*Black	4.382* (2.530)	0.0202 (0.0144)	4.746 (7.207)	-0.00394 (0.0236)	1.433 (8.284)	-0.0162 (0.0198)	0.983* (0.538)
Post Booker Judge*Hispanic	-0.677 (1.618)	0.0157 (0.0227)	-6.538 (8.008)	0.000139 (0.0306)	-2.496 (5.541)	-0.0159 (0.0331)	0.822** (0.366)
Post Booker Judge*Other	0.561 (2.654)	0.0451 (0.0310)	-12.61 (12.44)	0.0253 (0.0646)	-1.288 (6.763)	-0.0704 (0.0615)	2.613 (2.328)
Booker	-2.555** (1.254)	-0.000691 (0.0118)	1.446 (6.244)	0.0719*** (0.0224)	0.736 (4.371)	-0.0712*** (0.0231)	-0.401 (0.397)
Booker*Black	3.050*** (1.147)	0.0117** (0.00556)	15.03** (6.023)	-0.0134 (0.0109)	-3.143 (8.627)	0.00170 (0.0101)	0.823** (0.389)
Booker*Hispanic	0.760 (0.831)	-0.00167 (0.00569)	-0.227 (4.147)	-0.00486 (0.0173)	2.081 (4.288)	0.00653 (0.0156)	0.358* (0.204)
Booker*Other	0.0750 (1.103)	0.00327 (0.0116)	8.481 (12.74)	-0.00508 (0.0286)	0.0415 (5.458)	0.00182 (0.0245)	-0.954 (0.610)
Female Judge	1.418 (1.027)	0.00285 (0.00642)	-3.574 (6.059)	-0.0110 (0.0135)	3.612 (4.705)	0.00814 (0.0161)	0.328 (0.292)
Female Judge*Black	0.230 (1.602)	0.00242 (0.00911)	7.303 (6.260)	0.0145 (0.0180)	0.445 (12.81)	-0.0169 (0.0201)	-0.176 (0.368)
Female Judge*Hispanic	-0.996 (1.214)	0.00206 (0.00811)	3.262 (6.608)	0.000581 (0.0190)	2.271 (5.069)	-0.00264 (0.0211)	-0.333 (0.245)
Female Judge*Other	-2.031 (1.329)	-0.0124 (0.00963)	2.331 (9.701)	0.00735 (0.0264)	-10.92 (7.838)	0.00501 (0.0276)	-0.875 (0.566)
Female Judge*Booker	-2.417** (0.951)	-0.00351 (0.00879)	2.032 (6.990)	0.0208 (0.0142)	3.671 (4.586)	-0.0173 (0.0181)	-0.226 (0.380)
Female Judge*Booker*Black	-0.0853 (1.643)	-0.00367 (0.0121)	-10.97 (7.743)	-0.0221 (0.0214)	-7.574 (12.88)	0.0257 (0.0256)	-0.333 (0.452)
Female Judge*Booker*Hispanic	1.530 (0.982)	-0.000806 (0.0105)	0.134 (7.266)	-0.00556 (0.0161)	-9.481* (5.520)	0.00637 (0.0202)	0.372 (0.349)
Female Judge*Booker*Other	1.599 (2.263)	0.0127 (0.0176)	8.924 (15.01)	-0.0137 (0.0316)	7.373 (9.548)	0.00104 (0.0304)	0.325 (0.734)
Black Judge	-0.331 (0.793)	-0.00651 (0.00758)	-5.803 (4.309)	-0.0142 (0.0127)	-9.873 (11.31)	0.0207 (0.0158)	0.0118 (0.217)
Black Judge*Black	1.445 (1.322)	0.00771 (0.00815)	13.53** (5.807)	0.00716 (0.0137)	-11.16 (21.36)	-0.0149 (0.0164)	0.245 (0.327)
Black Judge*Hispanic	-0.405 (1.013)	-0.00252 (0.00967)	1.355 (5.835)	0.0378 (0.0241)	16.00 (11.44)	-0.0353 (0.0275)	0.0286 (0.296)
Black Judge*Other	1.587 (1.783)	-0.00655 (0.0147)	23.21* (11.89)	-0.0576** (0.0269)	2.950 (12.03)	0.0641* (0.0335)	-0.0150 (0.594)

TABLE 2.20. SENTENCING PATTERNS BEFORE AND AFTER *Booker*, BY JUDICIAL DEMOGRAPHICS (CONTINUED)

Black Judge*Booker	-0.615 (1.278)	0.00244 (0.0103)	6.650 (4.678)	0.0269 (0.0168)	10.27 (10.52)	-0.0294 (0.0208)	-0.0790 (0.317)
Black Judge*Booker*Black	-1.916 (1.669)	-0.00161 (0.00974)	-15.77** (6.071)	-0.00492 (0.0205)	11.38 (21.02)	0.00653 (0.0226)	-0.361 (0.503)
Black Judge*Booker*Hispanic	-0.543 (1.256)	0.00172 (0.0112)	-4.824 (6.605)	-0.0388 (0.0266)	-16.28 (10.63)	0.0371 (0.0284)	-0.224 (0.414)
Black Judge*Booker*Other	-1.470 (1.890)	0.0198 (0.0188)	-21.87* (13.01)	0.107*** (0.0345)	-8.142 (12.33)	-0.127*** (0.0401)	0.433 (0.746)
Democratic Judge	-0.969 (0.820)	0.00433 (0.00488)	-0.903 (4.072)	0.0210* (0.0109)	4.061 (5.135)	-0.0254** (0.0112)	-0.418* (0.218)
Democratic Judge*Black	-0.658 (1.055)	-0.00890 (0.00559)	6.518 (5.918)	0.0101 (0.00992)	-6.783 (9.201)	-0.00123 (0.0111)	-0.382 (0.296)
Democratic Judge*Hispanic	0.526 (1.007)	-0.00411 (0.00555)	1.370 (4.969)	-0.00607 (0.0125)	-7.242 (5.873)	0.0102 (0.0114)	0.218 (0.184)
Democratic Judge*Other	0.632 (1.218)	0.00936 (0.00895)	3.612 (8.762)	0.0129 (0.0246)	-0.615 (6.925)	-0.0223 (0.0239)	-0.799* (0.404)
Democratic Judge*Booker	0.176 (1.101)	-0.00773 (0.00694)	0.180 (4.848)	0.0142 (0.0134)	-3.031 (5.629)	-0.00642 (0.0146)	0.447* (0.246)
Democratic Judge*Booker*Black	-1.723 (1.410)	0.00920 (0.00748)	-10.39 (6.716)	0.00429 (0.0150)	8.125 (8.599)	-0.0135 (0.0157)	-0.154 (0.411)
Democratic Judge*Booker*Hispanic	-0.421 (1.142)	0.00878 (0.00741)	-0.857 (5.687)	-0.00946 (0.0140)	6.917 (6.440)	0.000684 (0.0141)	-0.261 (0.239)
Democratic Judge*Booker*Other	-0.647 (1.785)	0.000605 (0.0113)	-2.612 (13.02)	-0.0285 (0.0327)	0.319 (10.31)	0.0279 (0.0338)	0.194 (0.647)
Pre Guidelines Judge	-0.546 (0.761)	-0.00326 (0.00499)	-4.842 (4.541)	0.00841 (0.0125)	3.989 (9.640)	-0.00515 (0.0127)	0.206 (0.203)
Pre Guidelines Judge*Black	1.575 (1.260)	-0.00139 (0.00632)	12.08 (7.558)	-0.00217 (0.0113)	-26.72 (20.37)	0.00356 (0.0124)	0.0770 (0.351)
Pre Guidelines Judge*Hispanic	0.138 (0.791)	0.00254 (0.00490)	3.825 (5.218)	-0.0147 (0.0166)	-5.434 (9.774)	0.0121 (0.0170)	-0.210 (0.191)
Pre Guidelines Judge*Other	0.421 (1.349)	0.0256* (0.0140)	4.107 (9.500)	-0.0133 (0.0251)	-8.988 (13.36)	-0.0123 (0.0267)	-1.301** (0.531)
Pre Guidelines Judge*Booker	0.645 (1.197)	0.0132 (0.00931)	2.289 (5.505)	0.00772 (0.0157)	-1.550 (9.094)	-0.0209 (0.0146)	-0.277 (0.292)
Pre Guidelines Judge*Booker*Black	-2.228 (1.741)	-0.00519 (0.0114)	-10.61 (8.665)	0.00535 (0.0184)	22.34 (18.68)	-0.000152 (0.0171)	-0.300 (0.489)
Pre Guidelines Judge*Booker*Hispanic	-0.0513 (1.100)	-0.0108 (0.00869)	3.226 (7.071)	-0.00771 (0.0179)	2.848 (9.145)	0.0185 (0.0161)	0.346 (0.286)
Pre Guidelines Judge*Booker*Other	2.699 (2.648)	-0.0129 (0.0236)	-14.23 (15.35)	-0.0461 (0.0351)	5.441 (13.52)	0.0591 (0.0369)	1.052 (0.984)
Observations	214,136	214,136	13,091	214,136	82,432	214,136	118,612
R-squared	0.784	0.194	0.369	0.245	0.919	0.203	0.985

Notes: Data is from the matched USSC, TRAC, Federal Judicial Center data from 2000-2009 for courts with random assignment, excluding judges who formally retired prior to 2000. All regressions contain controls for offense type, and dummies for each offense level and criminal history combination. Regressions also contain district by sentencing year, and sentencing month fixed effects, and standard errors are clustered at the district level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

2.6.6. Response of Prosecutors to Increased Judicial Discretion

While the disparities estimated in this paper do not capture the compounded disparities that can result at each stage of the criminal process, I conclude by exploring the impact of increased judicial discretion on changes in prosecutorial decisions to charge mandatory minimums. Given that *Booker* left Congressionally-enacted statutory minimums intact, one would not necessarily expect judicial treatment of mandatory minimums to change in the aftermath of *Booker*. However, prosecutors may strategically respond to increased judicial discretion post *Booker* if they want to bind judges from departing downwards. After *Booker*, prosecutors have commented that they are far less willing to forego charging mandatory minimums because judges ultimately sentence defendants below the Guidelines minimum.

Table 2.21 presents results suggesting that prosecutorial discretion post *Booker* has not differentially affected black and white defendants in terms of charging offenses that carry mandatory minimums, although black and Hispanic defendants are far more likely to receive a mandatory minimum. However, black defendants are significantly more likely to face a *binding* mandatory minimum post *Booker* compared to white defendants.²⁶ The greater prevalence of binding mandatory minimums for black defendants in the aftermath of *Booker* suggests that more statutory minimums are applied to black defendants which exceed the Guidelines recommended sentences compared to similar white offenders. This finding suggests that black defendants may face statutory minimums that are harsher than the severity of the crime dictates, potentially indicating prosecutorial disparities post *Booker*.²⁷

²⁶This finding is robust to looking only at drug statutory minimums (the majority of statutory minimums cases) and controlling for specific drug type.

²⁷Main results from Table 2.6 are robust to controlling for the length of mandatory minimums rather than just a binary indicator). While the main results from Table 2.6 are robust to looking only at non mandatory minimum cases, point estimates are smaller in magnitude, suggesting that prosecutorial charging plays a role in exacerbating racial disparities in sentencing outcomes.

TABLE 2.21. TREATMENT OF MANDATORY MINIMUMS

	(1)	(2)	(3)	(4)
	Mandatory Minimum	Binding Minimum	Safety Valve	Substantial Assistance
Booker*Black	-0.00320 (0.00722)	0.0240*** (0.00870)	0.0146 (0.00938)	-0.0132 (0.0113)
Booker*Other	-0.0174*** (0.00569)	-0.00262 (0.00703)	0.0207* (0.0116)	-0.00440 (0.0122)
Booker*Hispanic	-0.0231** (0.0112)	-0.0162 (0.0192)	0.0268 (0.0322)	0.0507 (0.0308)
Black	0.0490*** (0.00745)	0.00678 (0.00673)	-0.0193*** (0.00601)	-0.0851*** (0.0107)
Hispanic	0.0477*** (0.00543)	0.0240*** (0.00536)	0.00275 (0.00940)	-0.0858*** (0.0101)
Other	-0.00959 (0.00800)	0.00393 (0.0130)	-0.0145 (0.0192)	-0.0613** (0.0292)
Booker	0.00158 (0.00830)	-0.0349** (0.0168)	-0.0119 (0.0151)	0.0299 (0.0225)
Race Trends?	Yes	Yes	Yes	Yes
Observations	816,564	244,273	162,294	221,320
R-squared	0.649	0.638	0.668	0.171

Notes: Data is from the USSC from 1994-2009. All regressions contain controls for offense type, and dummies for each offense level and criminal history combination. Regressions also contain district by sentencing year, and sentencing month fixed effects, and standard errors are clustered at the district level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level. When the dependent variable is safety valve, data is from 1999-2009.

However, conditional on being convicted of a charge that carries a mandatory minimum, decisions to reduce sentences below the mandatory minimum do not differ significantly by defendant race after *Booker*. A judge has some leeway in reducing sentence length for certain drug trafficking offenses under the “safety valve” provision under 18 U.S.C. §3553(f), which allows a judge to reduce the punishment for low level, first time offenders. Prosecutors also have the ability to reduce sentences below the mandatory minimum if the defendant offers “substantial assistance” during another investigation or prosecution under 18 U.S.C. §3553(e).

Column 3 suggest that the application of the safety valve does not change differentially post *Booker*, although black defendants are significantly less likely to receive the safety valve compared to similar white offenders for drug trafficking crimes.²⁸ Similarly, column 4 indicates that government sponsored substantial assistance motions for cases with mandatory minimums do not change differentially between offenders post *Booker*, although non white defendants are generally significantly less likely to receive substantial assistance motions.

2.7. Conclusion

After almost two decades of mandatory Guidelines sentencing, the Supreme Court struck down the Guidelines in *United States v. Booker*, greatly increasing the degree of judicial discretion. In subsequent decisions, the Court further increased judicial discretion by reducing the degree of appellate review and granting judges explicit permission to reject the policies of the Sentencing Commission.

Using comprehensive data on federal defendants sentenced from 1994-2009, I find evidence that increased judicial

²⁸This finding is also reported in the Sentencing Commission Report (2011) which states that in recent years, white defendants in drug cases are more frequently granted the safety valve exception compared to other defendants.

discretion via *Booker* has led to large and robust increases in racial disparities in sentencing, particularly after periods of reduced appellate scrutiny. By four years after *Booker*, the racial sentencing gap increases to 4.4 months, almost a 10% increase in the average sentence length. I also find that recent increases in racial disparities in sentencing appear to be driven by judges appointed post *Booker*, consistent with a story in which judges experienced with sentencing under rule-based sentencing continue to follow the Guidelines even when given more discretion. These findings should, however, be interpreted cautiously as they only apply to new George W. Bush appointees. Barack Obama appointed judges may exhibit different sentencing patterns. Finally, my results suggest that prosecutors charge black defendants with higher rates of binding mandatory minimums compared to white defendants after *Booker*, consistent with prosecutors attempting to bind judges to prevent them from departing downwards from the Guidelines in response to increased judicial discretion.

Despite the increase in racial disparities in federal sentencing after *Booker*, 75% of federal district judges believe that the current advisory regime better achieves the purposes of sentencing compared to the mandatory Guidelines regime prior to *Booker* (3%) or the “free at last” regime before the implementation of the Guidelines (8%). Only 14% of judges believe that a new mandatory Guidelines regime that complies with the Sixth Amendment would best achieve sentencing goals.

The findings in this paper suggest that while most federal district judges prefer the expanded judicial discretion under the current advisory system to the mandatory Guidelines regime, discretion comes with potentially undesirable consequences. An increase in disparities in the wake of increased judicial discretion can reflect unwarranted disparities if judicial bias enters into decision-making. On the other hand, disparities may be warranted if expanded discretion allows judges to tailor a sentence to the unique circumstances of an offender. For instance, disparities may emerge if judges are sentencing according to defendant characteristics, both observed and unobserved, that are correlated with actual recidivism risk.

In fact, recidivism rates are higher among nonwhite offenders, offenders with more extensive criminal histories and lower educational attainment, and I find that judges sentence these defendants to longer prison terms after *Booker*. Unconditional on other characteristics, black offenders are more likely to recidivate (32.8%) than Hispanic offenders (24.3%) and white offenders (16.0%) (United States Sentencing Commission 2004). Even controlling for basic demographics, criminal history and severity of offense, blacks are about 3.2 percentage points more likely to recidivate than white offenders (Kuziemko 2011). Taken with the finding that an additional month of time served reduces three-year recidivism by 1.3 percentage points, a judge would sentence black defendants to an additional 2.4 months in prison to equalize the recidivism rate across observably similar black and white defendants. This magnitude is consistent with the size of racial disparities I find in the aftermath of *Booker*, suggesting that increased disparities may be somewhat attributable to socially optimal sentencing aimed at reducing recidivism.

On the other hand, recidivism also varies greatly by gender and age after controlling for various observables, and

judges are unresponsive to these variables in the aftermath of *Booker*, indicating that judges are not solely sentencing based on actual recidivism risk. Future work could analyze the extent to which disparities in sentencing are warranted by looking at rates of recidivism in the federal criminal justice system. More generally, the framework in this paper can be applied to analyzing the impact of increased discretion on many other actors in the criminal justice system. Further work on the interactions of actors at various stages in the criminal process is critical to a thorough exploration of disparities in the federal criminal justice system.

3. HAVE INTER-JUDGE SENTENCING DISPARITIES INCREASED IN AN ADVISORY GUIDELINES REGIME? EVIDENCE FROM *BOOKER*

3.1. Introduction

The Federal Sentencing Guidelines were adopted to counter widespread disparities in federal sentencing. By the 1970s, the federal system exhibited “an unjustifiably wide range of sentences to offenders convicted of similar crimes” because each judge was “left to apply his own notions of the purposes of sentencing.”²⁹ Disparities were so pronounced that a defendant sentenced to three years by one judge would have been sentenced to twenty years had he been assigned to another judge.³⁰ Decrying these disparities and championing sentencing reform, federal district judge Marvin Frankel claimed that “the almost wholly unchecked and sweeping powers we give to judges in the fashioning of sentences are terrifying and intolerable for a society that professes devotion to the rule of law.”³¹

In response, policymakers sought to limit the “unfettered discretion the law confers on those judges and parole authorities [that implement] the sentence.”³² Under the Sentencing Reform Act (SRA) of 1984, Congress created the United States Sentencing Commission to promulgate the Guidelines,³³ a new regime intended to reduce disparities stemming from judicial preferences and biases rather than offense and offender characteristics.³⁴ Congress directed the Commission, an independent agency within the judicial branch,³⁵ to fashion sentencing guidelines aimed at the primary goal of avoiding unwarranted sentencing disparity.³⁶

After the implementation of the Guidelines, researchers began to investigate the extent to which the Guidelines reduced disparities.³⁷ Initial work by Anderson, Stith, and Kling revealed that the Guidelines was somewhat successful

²⁹See S. REP. NO. 97-307 at 5 (1981).

³⁰Anthony Partridge & William B. Eldridge, Federal Judicial Center, *The Second Circuit Study: A Report to the Judges of the Second Circuit* 36 (1974).

³¹Marvin E. Frankel, *CRIMINAL SENTENCES: LAW WITHOUT ORDER* 5 (1973).

³²See S. Rep. No. 98-225 at 38 (1983) (Senate Report on precursor to federal Sentencing Reform Act of 1984) (“[E]very day Federal judges mete out an unjustifiably wide range of sentences to offenders with similar histories, convicted of similar crimes, committed under similar circumstances. . . . These disparities, whether they occur at the time of the initial sentencing or at the parole stage, can be traced directly to the unfettered discretion the law confers on those judges and parole authorities responsible for imposing and implementing the sentence”); *Id.* at 49 (“[T]he present practices of the federal courts and of the parole commission clearly indicate that sentencing in the federal courts is characterized by unwarranted disparity and by uncertainty about the length of time offenders will service in prison.”).

³³Pub. L. No. 98-473, 98 Stat 1837, 1987 (1984) (codified as amended at 18 USC §3551 et seq., 28 USC §991 et seq.). For a discussion of the efforts leading up to the promulgation of the SRA, see Kate Stith & Steve Y. Koh, *The Politics of Sentencing Reform: The Legislative History of the Federal Sentencing Guidelines*, 28 WAKE FOREST L. REV. 223, 230 (1993).

³⁴U.S.S.G. §1A.1, intro to comment., pt. A, ¶2 (Congress “sought uniformity in sentencing by narrowing the wide disparity in sentences imposed by different federal courts for similar criminal conduct”); S. REP. NO. 225 at 45 (1983) (“Sentencing disparities that are not justified by differences among offenses or offenders are unfair both to offenders and to the public.”).

³⁵See 28 U.S.C. §991(a). The Commission was placed in the Judicial Branch because Congress concluded that “sentencing should remain primarily a judicial function,” and because sitting judges would serve on the Commission. The Commission is comprised of seven voting members. The SRA provides that “[a]t least three of the [Commission] members shall be Federal judges selected after considering a list of six judges recommended to the President by the Judicial Conference of the United States” and no more than four members of the Commission could be members of the same political party. See *id.* The Commission later withstood separation-of-powers challenges, as the Court rejected several constitutional challenges to the Commission and its delegated authority. See *Mistretta v. United States*, 488 U.S. 361 (1989).

³⁶See, e.g. Stephen Breyer, *The Federal Sentencing Guidelines and the Key Compromises Upon Which They Rest*, 17 HOFSTRA L. REV. 1, 3 (1988) (Congress sought to reduce “unjustifiably wide” sentencing disparity.); Kenneth R. Feinberg, *Federal Criminal Sentencing Reform: Congress and the United States Sentencing Commission*, 28 WAKE FOREST L. REV. 291, 295 (1993) (“The first and foremost goal of the sentencing reform effort was to alleviate the perceived problem of federal criminal sentencing disparity.”).

³⁷While the Guidelines were effectively mandatory, the Guidelines did provide a permissible range of sentence lengths. Thus, one should expect that the Guidelines would ameliorate, but not completely eliminate, inter-judge disparities.

in reducing inter-judge sentencing disparity.³⁸ The authors estimate that the expected difference in sentence length between judges fell from 17% (4.9 months) in 1986-1987 to 11% (3.9 months) in 1988-93 in 25 cities where assignment was found sufficiently random.³⁹ Another study by Hofer, Blackwell, and Ruback also found evidence of reduced inter-judge sentencing disparities after the promulgation of the Guidelines.⁴⁰ The study concludes that the Guidelines achieved “modest success” in reducing inter-judge disparities, documenting that the sentencing judge accounted for 2.32% of the variation in sentences in 1984-1985, but only 1.24% under the Guidelines in 1994-1995.⁴¹ Convergence in findings by outside researchers and the Commission led the Commission to conclude that “the federal sentencing guidelines have made significant progress toward reducing disparity caused by judicial discretion.”⁴²

After almost two decades of mandatory Guidelines sentencing, the Guidelines were struck down in *United States v. Booker*,⁴³ dramatically altering the sentencing landscape. In *Booker*, the Supreme Court found that the Guidelines violated the Sixth Amendment right to a jury right and rendered the Guidelines advisory. Subsequent Supreme Court decisions furthered weakened the effect of the Guidelines on criminal sentencing by reducing the degree of appellate scrutiny for both within and outside Guidelines sentences. In *Rita v. United States*, the Court directed court of appeals to apply a presumption of reasonableness to within Guidelines sentencing.⁴⁴ Later in *Gall v. United States*, the Court held that appellate courts could not presume that a sentence outside the Guidelines range was unreasonable, reducing the degree of appellate review to a more deferential abuse of discretion standard.⁴⁵ Concurrent with the *Gall* decision, the Court in *Kimbrough* held that federal district court judges have the discretion to impose sentences outside the recommended Guidelines range due to policy disagreements with the Sentencing Commission.⁴⁶

Following *Booker* and its progeny, *Rita*, *Gall*, *Kimbrough*, the legal community expressed concerns on the impact of such decisions on inter-judge sentencing disparities. Congressman Tom Feeney wrote that “the Supreme’s Court decision [in *Booker*] to place this extraordinary power to sentence a person solely in the hands of a single federal judge - who is accountable to no one - flies in the face of the clear will of Congress.”⁴⁷ U.S. Attorney for the Northern District of Illinois, Patrick Fitzgerald, stated that *Booker* has “re-introduced into federal sentencing both substantial district-to-district variations and substantial judge-to-judge variations.”⁴⁸ Similarly, scholars commented on the huge

³⁸James M. Anderson et al., *Measuring Interjudge Sentencing Disparity: Before and After the Federal Sentencing Guidelines*, 42 J.L. & ECON. 271, 303 (1999) (“The Guidelines have reduced the net variation in sentence attributable to the happenstance of the identity of the sentencing judge.”).

³⁹*Id.* at 294.

⁴⁰Paul J. Hofer, Kevin R. Blackwell, & R. Barry Ruback, *The Effect of the Federal Sentencing Guidelines on Inter-Judge Sentencing Disparity*, 90 J. CRIM. L. & CRIMINOLOGY 239 (1999). *See id.* at 284-86 for a discussion of the statistical techniques employed in the study.

⁴¹*Id.* at 241, 287 (“Together with the other research reviewed below, [our] findings suggest that the sentencing guidelines have had modest but meaningful success at reducing unwarranted disparity among judges in the sentences imposed on similar crimes and offenders.”)

⁴²United States Sentencing Commission, FIFTEEN YEARS OF GUIDELINES SENTENCING: AN ASSESSMENT OF HOW WELL THE FEDERAL CRIMINAL JUSTICE SYSTEM IS ACHIEVING THE GOALS OF SENTENCING REFORM (Nov. 2004) at 99.

⁴³543 U.S. 220 (2005).

⁴⁴551 U.S. 338, 347-50 (2007).

⁴⁵552 U.S. 38, 52-53 (2007).

⁴⁶*Kimrough v. United States*, 552 U.S. 85, 101 (2007).

⁴⁷Carl Hulse & Adam Liptak, *New Fight Over Controlling Punishments Is Widely Seen*, N.Y. TIMES, Jan. 13, 2005, at A29.

⁴⁸Patrick J. Fitzgerald, Testimony before the U.S. Sentencing Commission Regional Hearing in Chicago, Illinois, at 3 (Sept. 10, 2009).

increase in the degree of judicial discretion afforded to judges,⁴⁹ and predicted an increase in unwarranted sentencing disparities.⁵⁰

Due to suggestive evidence of increasing disparities post *Booker*, the United States Sentencing Commission and policymakers have commented on possible ways to constrain judicial discretion. Then Attorney General Alberto Gonzales claimed that in light of “increasing disparity in sentences” since *Booker*, the Guidelines needed to be fixed.⁵¹ As a potential “fix,” former Chair of the Sentencing Commission, Judge William K. Sessions III, has suggested “resurrecting” the mandatory Guidelines in order to give them greater weight during sentencing.⁵²

On the other side of the debate, some scholars have suggested that *Booker* improved the “quality, transparency, and rationality” in federal sentencing, and thus *Booker* is the “fix.”⁵³ Indeed, the vast majority of federal district court judges also indicate that they prefer the current advisory guidelines system to alternative regimes. In a 2010 Sentencing Commission survey of district court judges, 75% indicated that the current advisory guidelines system “best achieves the purposes of sentencing,” while only 3% preferred the mandatory guidelines regime in place before *Booker*.⁵⁴

This Article asks the question: What has been the result of reintroduction of greater judicial discretion after *Booker* on inter-judge disparities? The primary empirical work on the impact of *Booker* on sentencing disparities is suggestive of increases in inter-judge sentencing disparity. Using sentencing data from the District of Massachusetts, Scott observes an increase in judge differences.⁵⁵ While a first step, the study is limited to ten judges in the Boston courthouse.

⁴⁹ See Nancy Gertner, *A Short History of American Sentencing: Too Little Law, Too Much Law, or Just Right*, 100 J. CRIM. L. & CRIMINOLOGY 691, 706 (2010) (“It is clear that *Booker* has enhanced the position of the judge, whose sentencing expertise has been formally acknowledged again, at the cost of diminishing the position of the Sentencing Commission.”); Douglas A. Berman, *Foreword: Beyond Blakely and Booker: Pondering Modern Sentencing Process*, 95 J. CRIM. L. & CRIMINOLOGY 653, 676 (2005) (“*Booker* devised a new system of federal sentencing which granted judges more sentencing power than they had ever previously wielded”); Luiza Ch. Savage, *Chaos Ahead After Sentencing Guidelines Decision*, N.Y. SUN, Jan. 13, 2005, at 1 (quoting Frank Bowman) (arguing that *Booker* resulted in “the most amount of judicial discretion ever afforded to sentencing judges”).

⁵⁰ See Bowman, *Debacle: How the Supreme Court Has Mangled American Sentencing Law and How It Might Yet Be Mended*, 77 U. CHI. L. REV. X, X (2010) (“*Kimbrough*, *Gall* ... have so thoroughly denatured appellate review that the federal system’s ability to control regional and judge-to-judge sentencing disparity has been effectively eliminated”).

⁵¹ Federal Sentencing Guidelines Speech (June 21, 2005), 17 FED. SENTENCING REP. 324, 325-26 (2005).

⁵² William K. Sessions III, *At the Crossroads of the Three Branches: The U.S. Sentencing Commission’s Attempts to Achieve Sentencing Reform in the Midst of Inter-Branch Power Struggles*, 26 J.L. & POL. 305, 346-50 (2011).

⁵³ See Amy Baron-Evans & Kate Stith, *Booker Rules*, 160 U. PA. L. REV. 1631, 1633 (2012).

⁵⁴ U.S. Sentencing Commission, RESULTS OF SURVEY OF UNITED STATES DISTRICT JUDGES JANUARY 2010 THROUGH MARCH 2010 (June 2010) (Question 19, Table 19).

⁵⁵ See Ryan W. Scott, *Inter-judge Sentencing Disparity After Booker: A First Look*, 63 STAN. L. REV. 1, 4-5 (2010). Scott finds almost a doubling of the effect of judge on sentence length post *Booker*, resulting in an average difference of over two years between lenient and harsh judges for cases not subject to a mandatory minimum. *Id.* at 40-41. Scott also finds significant variation in the rate of below-range sentencing among judges. Some judges sentenced below-range at the same rate prior to *Booker* (around 16%), while others increased their rate of sentencing below-range to as high as 53%. *Id.* The Transactional Records Access Clearinghouse (TRAC) recently compiled a dataset of the sentencing records of over 800 federal judges from fiscal year 2007 to 2011. See Susan B. Long & David Burnham, *TRAC Report: Examining Current Federal Sentencing Practices: A National Study of Differences Among Judges*, 25 FED. SENTENCING REP. 6, 7 (2012); see also *Big Sentencing Disparity Seen for Judges*, N.Y. TIMES, March 5, 2012, at A23. Relying on the random assignment of cases to judges within district courthouses, the TRAC study found statistically significant, unexplained differences in the typical sentences of judges in over half of the courthouses studied. *Id.* at 15. The most recent Commission report (2012) also finds suggestive evidence that variation among judges within the same district, in particular the rates of non-government sponsored below range sentences, has increased after *Booker* and *Gall*. United States Sentencing Commission, REPORT ON THE CONTINUING IMPACT OF UNITED STATES V. BOOKER ON FEDERAL SENTENCING (2012). The Commission concludes that “sentencing outcomes increasingly depend upon the judge to whom the case is assigned.” *Id.* at X. However, the Commission does not account for caseload composition differences across judges within the same district, and analyzes all 94 districts, despite evidence by previous researchers that random assignment of cases is not universal. Thus, the Commission’s findings are only suggestive.

Therefore, a comprehensive analysis of disparities post *Booker* is essential to informing these policy debates.⁵⁶ This Article fills this gap by undertaking the first large-scale, multi-district analysis of inter-judge sentencing disparities in federal sentencing after *Booker* by utilizing a new and comprehensive dataset constructed for this study. The Article proceeds in five parts. Part 3.2 provides a brief background of the legal landscape. Part 3.3 describes the dataset and empirical methods. Matching three data sources, I construct a dataset of over 600,000 criminal defendants linked to sentencing judge from 2000-2009.

Part 3.4 presents empirical results. Relying on the random assignment of cases to judges in district courthouses, I find evidence of significant increases in inter-judge disparities. A defendant who is randomly assigned a “harsher” judge in the district court received a 2.6 month longer prison sentence before *Booker*, but received a 5.3 month longer sentence following *Kimbrough/Gall*. Similarly, a defendant randomly assigned to a more “lenient” judge faced a 4.8% chance of receiving a below range departure before *Booker*, but over 6.6% chance after *Kimbrough/Gall*.

Part 3.4.4 undertakes an analysis of the sources of increases in inter-judge disparities. Many scholars have suggested that judges have different sentencing philosophies,⁵⁷ which may be affected by the standard of appellate review.⁵⁸ Sentencing practices are correlated with judge demographic characteristics such as race,⁵⁹ gender,⁶⁰ and political affiliation.⁶¹ In particular, the inevitable shift in the composition of the federal district courts may have profound

⁵⁶ Another strand of empirical research analyzes the impact of *Booker* on racial disparities in sentencing. The United States Sentencing Commission has found evidence of large racial disparities in sentencing outcomes after *Booker*. See United States Sentencing Commission, DEMOGRAPHIC DIFFERENCES IN FEDERAL SENTENCING PRACTICES: AN UPDATE OF THE BOOKER REPORT’S MULTIVARIATE REGRESSION ANALYSIS (2010) (providing evidence demographic differences were significantly less when the guidelines were binding, particularly during the PROTECT Act when appellate review of departures involved de novo review); United States Sentencing Commission, FINAL REPORT ON THE IMPACT OF UNITED STATES V. BOOKER ON FEDERAL SENTENCING, at 105-108 (2006). However, other scholars have no significant change in racial disparities, at least in sentence length. See J.T. Ulmer et. al, *Racial Disparity in the Wake of the Booker/Fanfan Decision: An Alternative Analysis to the USSC’s 2010 Report*, 10 CRIM. & PUB. POL’Y 1077 (2011). Some scholars argue that judicial discretion may actually mitigate recent increases in racial disparities. See Joshua Fischman & Max Schanzenbach, *Racial Disparities under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums*, J. EMPIRICAL LEGAL STUD. (forthcoming) (arguing that recent increases in racial disparities are mainly due to the increased relevance of mandatory minimums).

⁵⁷ See John S. Carroll et al., *Sentencing Goals, Casual Attributions, Ideology, and Personality*, 52 J. PERSONALITY & SOC. PSYCHOL. 107 (1987) (arguing that judicial ideology is reflected in how a judge thinks about the causes of crime and the goals of sentencing); Shari S. Diamond & Hans Zeisel, *Sentencing Councils: A Study of Sentence Disparity and Its Reduction*, 43 U. CHI. L. REV. 109, 114 (1975) (“it is reasonable to infer that the judges’ differing sentencing philosophies are a primary cause of the disparity”); see also, Paul Hofer, Kevin Blackwell, & R. Barry Ruback, *The Effect of the Federal Sentencing Guidelines on Inter-Judge Sentencing Disparity*, 90 J. CRIM. L. & CRIMINOLOGY 239 (1999) (claiming that there are differences between how liberals and conservatives view the goals of sentencing which can drive different sentencing practices).

⁵⁸ Joshua Fischman & Max Schanzenbach, *Do Standards of Review Matter? The Case of Federal Criminal Sentencing*, 40 J. LEG. STUD. 405 (2011). The authors find that Democratic appointees are more lenient than Republican appointees and differences in sentencing practices increase when appellate review is more deferential, suggesting that judges are constrained by the fear of reversal. The authors also find evidence that pre-Guidelines appointed judges are more likely to depart from the Guidelines.

⁵⁹ See Thomas Uhlman, RACIAL JUSTICE: BLACK JUDGES AND DEFENDANTS IN AN URBAN TRIAL COURT (1979) (claiming that black and white judges sentenced black defendants more harshly compared to white defendants); Susan Welch et al., *Do Black Judges Make a Difference?*, 32 AM. J. POL. SCI. 126 (1988) (finding that black judges do not differ much in their incarceration decisions from white judges based on one city’s criminal cases).

⁶⁰ See, e.g., Darrell Steffensmeier & Chris Herbert, *Women and Men Policy Makers: Does the Judge’s Gender Affect the Sentencing of Criminal Defendants?*, 77 SOCIAL FORCES 1163 (1999) (female judges sentence defendants for longer terms, are more likely to incarcerate minorities, and less likely to incarcerate women, in Pennsylvania criminal cases); Max Schanzenbach, *Racial and Sex Disparities in Sentencing: The Effect of District-Level Judicial Demographics*, 34 J. LEG. STUD. 57 (2005) (some evidence that minority and female judges sentence differently using district level variation in judicial characteristics).

⁶¹ For instance, Max M. Schanzenbach & Emerson H. Tiller, *Strategic Judging Under the United States Sentencing Guidelines: Positive Theory and Evidence*, 23 J.L. ECON. & ORG. 24 (2007) explore the impact of ideology on federal criminal sentencing decisions from 1992 through 2001. They find that sentences for serious crimes in districts comprised of more Democrat appointed judges are lower than sentences in districts with more Republican appointed judges. The alignment of the ideology of the reviewing court also increased departures from the Guidelines. More recent work in Max M. Schanzenbach & Emerson H. Tiller, *Reviewing the Sentencing Guidelines: Judicial Politics, Empirical Evidence, and Reform*, 75 U. CHI. L. REV. 715 (2008) reveals that Republican appointed judges give longer sentences for the same crime compared to their Democrat

consequences on unwarranted disparities as judges who have no experience sentencing under a presumptive guidelines regime take the federal bench.⁶² Federal defense lawyer James Felman predicted that following advisory guidelines, “unwarranted disparity in the near term would be considerably less than that which existed prior to 1987,” but “there will be a minority of judges who will generate unwarranted disparity, and this number seems likely to increase as the years go by and the bench is filled with individuals who have no history with binding guidelines.”⁶³

I find that female judges and Democratic appointed judges issue shorter sentences and are more likely to depart downwards than their male and Republican appointed peers, respectively. Also striking is the differential tendencies of post *Booker* and pre *Booker* judicial appointees. Judges who have no prior experience sentencing under the mandatory guidelines regime are more likely to depart from the guidelines recommended range than their pre-*Booker* counterparts, potentially suggesting that newer judges are less anchored to the Guidelines.

In addition to analyzing the impact of philosophy on sentencing practices of individual judges, this paper also contributes to the literature on geographical variations in sentencing patterns, which has found that court cultures can affect sentencing through both exercise of judicial discretion and also differences in policies among prosecutors in certain regions.⁶⁴ In Part 3.4.5, I present evidence of substantial inter-district differences in sentence length, below range departures, average rates of mandatory minimums, and average rates of substantial assistance motions.

In Part 3.4.6, I present some evidence on the contribution of prosecutorial decisions on inter-judge disparities. Undoubtedly, a defendant’s sentence is determined by the discretionary actions of multiple actors in the criminal justice process, culminating in sentencing. Therefore, any study of inter-judge sentencing disparities is only a partial portrayal of the disparities that can arise in the criminal justice system. Previous scholars rightly suggested that arrest, charge, and plea bargaining decisions made earlier in the process are all ripe avenues for unwarranted bias.⁶⁵

I find evidence that the application of a mandatory minimum is a large contributor of inter-judge and inter-district

appointed counterparts. Moreover, Democrat-appointed judges are more likely to depart downwards from the Guidelines when the reviewing circuit court is majority Democrat appointed.

⁶²Until now, prior studies have been unable to identify the impact of post *Booker* appointed judges on inter-judge sentencing disparities. The Scott study, which only looks at the Boston courthouse, is unable to take into account changes in judicial composition because the Boston courthouse did not experience turnover during the years in his study. Recent work suggest that racial disparities in sentencing are greater among judges appointed after *Booker*. See Crystal Yang, *Free At Last? Judicial Discretion and Racial Disparities in Federal Sentencing*, Olin Fellows’ Discussion Paper No. 47 (2012).

⁶³See James Felman, *How Should Congress Respond if the Supreme Court Strikes Down the Federal Sentencing Guidelines?*, 17 FED. SENTENCING REP. 97, 98-99 (2004).

⁶⁴See, e.g., Paul L. Sutton, Department of Justice, *Federal Sentencing Patterns: A Study of Geographical Variations* (1978); William M. Rhodes & Catherine Conly, Department of Justice Federal Justice Research Program, *Analysis of Federal Sentencing* (1981); Charles D. Weisselberg & Terrence Dunworth, *Inter-District Variation under the Guidelines: The Trees May Be More Significant Than the Forest*, 6 FED. SENTENCING REP. 25, 26-27 (1993) (finding that the guidelines do not impact all cases and all districts equally and that the guidelines mean different things in different contexts).

⁶⁵See, e.g., Franklin E. Zimring, *Making the Punishment Fit the Crime*, 6 HASTINGS CENTER REP. 13, 13-14 (1976) (arguing that there is “multiple discretion” in the criminal justice system from the legislature, prosecutor, judge and parole board); Ilene H. Nagel & Steven J. Schulhofer, *A Tale of Three Cities: An Empirical Study of Charging and Bargaining Practices Under the Sentencing Guidelines*, 66 S. CAL. L. REV. 501, 502 (1992) (noting that “both Congress and the U.S. Sentencing Commission were well aware that plea bargaining posed a potential threat to the success of guidelines sentencing”). As a result, the Guidelines system has been attacked by many for its rigidity and for shifting power to prosecutors in their charge and plea bargaining decisions. See, e.g., Daniel J. Freed, *Federal Sentencing in the Wake of the Guidelines: Unacceptable Limits on the Discretion of Sentencers*, 101 YALE L.J. 1681, 1719-20, 1725-27 (1992); Kate Stith, *The Arc of the Pendulum: Judges, Prosecutors, and the Exercise of Discretion*, 117 YALE L.J. 1420, 1430 (2008); See Albert W. Alschuler, *Sentencing Reform and Prosecutorial Power: A Critique of Recent Proposals for “Fixed” and “Presumptive” Sentencing*, 126 U. PA. L. REV. 550 (1978); Kate Stith & Jose A. Cabranes, *FEAR OF JUDGING: SENTENCING GUIDELINES IN THE FEDERAL COURTS* 1998.

disparities, such that measures of variation are reduced by almost a factor of two when I exclude cases in which mandatory minimums are charged. The results suggest substantial unequal application of mandatory minimums to similar cases within a district courthouses, and different mandatory minimum policies by prosecutors across district courts. There are also substantial differences in the rates of substantial assistance motions filed by prosecutors across judges and district courts.

In Part 3.5, I describe recent proposals to reform federal sentencing and apply the empirical findings in this paper to shed light on the soundness of the proposals. Given the finding that a substantial portion of inter-judge disparities and regional disparity is attributable to the application of mandatory minimums, any proposal that contemplates shifting power to prosecutors will likely exacerbate the presence of disparities. Indeed, many judges and scholars have suggested that mandatory minimums are “fundamentally inconsistent with the sentencing guidelines system.”⁶⁶ Instead, I argue that strengthened appellate review and elimination of mandatory minimums are potential steps in the direction of reducing unwarranted disparities in sentencing. Part 3.6 concludes.

3.2. Brief Legal Background of Federal Sentencing

3.2.1. Adoption of the Federal Sentencing Guidelines

In the early twentieth century, criminal justice was premised on the notion of rehabilitation.⁶⁷ This goal of rehabilitation manifested itself in the form of indeterminate sentencing, which allowed prison sentences and probation to be tailored to each offender’s progress toward rehabilitation. As a result, judges and parole boards possessed substantial discretion in their sentencing determinations.⁶⁸ In this regime of “free at last” sentencing,⁶⁹ federal judges had essentially unlimited authority in imposing sentences, limited only by statutorily prescribed minimum and maximum sentences.⁷⁰ Lack of appellate review of sentences meant that judges faced no meaningful check to ensure uniformity

⁶⁶See Sessions, *supra* note X, at 329 (citing Senator Edward M. Kennedy, *Sentencing Reform-An Evolutionary Process*, 3 FED. SENTENCING REP. 271, 272 (1991) (“Mandatory minimum sentencing statutes have . . . hampered the guideline system and are becoming an increasingly serious obstacle to its success. . . . Mandatory minimums inevitably lead to sentencing disparity because defendants with different degrees of guilt and different criminal records receive the same sentence.”)); Stephen Breyer, *Federal Sentencing Guidelines Revisited*, 11 FED. SENTENCING REP. 180, 184-85 (1999) (“ . . . Congress, in simultaneously requiring Guideline sentencing and mandatory minimum sentences, is riding two different horses. And those two horses, in terms of coherence, fairness, and effectiveness, are traveling in opposite directions. . . . [Congress needs to] abolish mandatory minimums altogether.”).

⁶⁷See David Rothman, *Sentencing Reform in Historical Perspective*, 29 CRIME & DELINQUENCY 631, 637-41 (1983) (reformers “pursue[d] rehabilitation, which meant treating the criminal not the crime, calculating the sentence to fit the individual needs and characteristics of the offender”); see also Frank O. Bowman, III, *The Quality of Mercy Must Be Restrained, and Other Lessons in Learning to Love the Federal Sentencing Guidelines*, 1996 WIS. L. REV. 679, 680-89 (describing the rehabilitative sentencing model).

⁶⁸See Rothman, *supra* note X at 638 (“the judge would make his decision (probation or such a minimum-maximum term); eventually the prison classification committee experts would make their decision (this program or that program), and the parole board experts would make theirs (release at the minimum, or later)”).

⁶⁹A term coined by Judge Nancy Gertner to describe the state of indeterminate sentencing prior to 1984. See *United States v. Mueffelman*, 400 F. Supp. 2d 368, 372 (D. Mass 2005); *United States v. Jaber*, 362 F. Supp. 2d 365, 370 (D. Mass. 2005).

⁷⁰See United States Sentencing Commission, *THE FEDERAL SENTENCING GUIDELINES: A REPORT ON THE OPERATION OF THE GUIDELINES SYSTEM AND SHORT-TERM IMPACTS ON DISPARITY IN SENTENCING, USE OF INCARCERATION, AND PROSECUTORIAL DISCRETION AND PLEA BARGAINING*, Vol. I (December 1991) at 9 (judges “decided the various goals of sentencing, the relevant aggravating and mitigating circumstances, and the way in which these factors would be combined in determining a specific sentence”); see *Koon v. United States*, 518 U.S. 81, 96 (1996) (“Before the Guidelines system, a federal criminal sentence within statutory limits was, for all practical purposes, not reviewable on appeal.”)

and consistency in sentencing.⁷¹

By the 1970s, faith in the rehabilitative model of sentencing declined due to a confluence of changing social norms, escalating public anxiety over rising crime rates, and public skepticism of the ability to rehabilitate criminal offenders.⁷² The legal community and public expressed alarm at the widespread disparities in criminal sentencing. Some argued that judges and parole boards endangered public safety with lenient sentencing of criminals.⁷³ Others were distressed by inequitable and arbitrary treatment within the criminal justice system as studies showed that similar offenders were often punished very differently. A 1977 study of 47 Virginia state district court judges revealed that while judges generally agreed on the verdict in legal cases, they applied radically different sentences.⁷⁴ A Federal Judicial Center Second Circuit Study found large inter-judge differences in the sentences imposed based on identical presentence reports of defendants.⁷⁵ The same defendant was sentenced to three years by one judge, and twenty years by another.⁷⁶ Some concluded that this disparate treatment of defendants by judges produced racial inequities in sentencing. The American Friends Service Committee claimed that decreasing discretion among judges and parole boards was the only way to eliminate racial discrimination and sentencing disparities in the criminal justice system.⁷⁷

Other studies identified large inter-court differences. A 1988 study of federal courts found that white collar offenders who committed similar offenses received very different sentences depending on the court in which they were sentenced,⁷⁸ with one study observing that “a defendant sentenced by a Southern judge was likely to serve six months more than average, while a defendant sentenced in Central California was likely to serve twelve months less.”⁷⁹

These large disparities in sentencing prompted calls for sentencing reform. Championing the call for reform, federal district judge Marvin Frankel expressed grave concern over the indeterminate and individualized sentencing regime of the period, claiming that “the almost wholly unchecked and sweeping powers we give to judges in the fashioning of sentences are terrifying and intolerable for a society that professes devotion to the rule of law.”⁸⁰ Judge Frankel called for the creation of an independent sentencing commission that would replace judicial and parole board

⁷¹ See, e.g., *Dorszynski v. United States*, 418 U.S. 424, 431 (1974) (stating the general proposition that appellate review ends if a sentence is within the limitations set forth in the statute).

⁷² See, e.g. Francis A. Allen, *THE DECLINE OF THE REHABILITATIVE IDEAL: PENAL POLICY AND SOCIAL PURPOSE*, 25-30 (1981); see also Frank O. Bowman, III, *Debacle: How the Supreme Court Has Mangled American Sentencing Law and How It Might Yet Be Mended*, 77 U. CHI. L. REV. X (forthcoming 2010) (attributing demand for social controls to rising crime rates and social upheaval).

⁷³ Michael Tonry, *Obsolescence and Immanence in Penal Theory and Policy*, 105 COLUM. L. REV. 1233, 1247 (2005) (conservatives criticized indeterminate sentencing for being uncertain and lenient and increasing crime rates).

⁷⁴ William Austin & Thomas A. Williams, III, *A Survey of Judges' Responses to Simulated Legal Cases: Research Note on Sentencing Disparity*, 69 J. CRIM. L. & CRIMINOLOGY 306 (1977).

⁷⁵ Anthony Partridge & William B. Eldridge, Federal Judicial Center, *The Second Circuit Study: A Report to the Judges of the Second Circuit* 36 (1974).

⁷⁶ *Id.*

⁷⁷ See Am. Friends Serv. Comm., *Struggle for Justice: A Report on Crime and Punishment in America* 130 (1971) (claiming that discretion allowed judges and parole boards to control minority groups); see also Bowman, *Quality of Mercy*, *supra* note X, at 686-88 (critics arguing discretion produced unjustifiable disparities open to conscious or unconscious racial and other biases, demanding “truth in sentencing”).

⁷⁸ Wheeler et al., *SITTING IN JUDGMENT: THE SENTENCING OF WHITE-COLLAR CRIMINALS* 1988.

⁷⁹ See Justice Stephen Breyer, *Federal Sentencing Guidelines Revisited*, 11 FED. SENTENCING REP. 180 (1988).

⁸⁰ Marvin E. Frankel, *CRIMINAL SENTENCES: LAW WITHOUT ORDER* 5 (1973) (Frankel also argued that individualized sentencing was “out of hand,” and criticized the state of indeterminate sentencing, at 10, 26-49.); see also Marvin E. Frankel, *Lawlessness in Sentencing*, 41 U. CIN. L. REV. 1 (1972).

discretion.⁸¹

In response, Congress created the United States Sentencing Commission to adopt and administer the Sentencing Guidelines, aimed at eliminating unwarranted sentencing disparities “among defendants with similar records who have been found guilty of similar criminal conduct.”⁸² Part of the SRA of 1984, the Guidelines were applied to all federal offenses committed after November 1, 1987. The SRA has been viewed by some as creating a regime that preserves some judicial discretion,⁸³ while others have viewed the SRA as substantially increasing the severity of punishment and dramatically reducing the discretion afforded to judges to consider the particular circumstances of each offender.⁸⁴

Notwithstanding disagreement about the degree to which sentencing reform changed the legal landscape, the new SRA introduced a shift from a regime of virtually unfettered judicial discretion to more restricted discretion within a system of determinate sentencing.⁸⁵ By requiring judges to sentence within the recommended Guidelines range unless the court found aggravating or mitigating circumstances,⁸⁶ the Guidelines were treated as presumptively mandatory, although the particular standards for departure were ambiguous.⁸⁷ Later in *Koon v. United States*, the Supreme Court held that a district court judge’s decision to depart from the Guidelines range would be subject to an abuse of discretion standard of appellate review.⁸⁸

Under the Guidelines, federal district court judges assign each federal crime to one of 43 offense levels, and assign each federal defendant to one of 6 criminal history categories. The more serious the offense and the greater the harm associated with the offense, the higher the base offense level assigned under Chapter Two of the Guidelines.⁸⁹ For example, trespass offenses are assigned a base offense level of 4,⁹⁰ while kidnapping is assigned a base offense level of 32.⁹¹ From a base offense level, the final offense level is calculated by adjusting for applicable offense characteristics

⁸¹Frankel, CRIMINAL SENTENCES: LAW WITHOUT ORDER (1973).

⁸²28 U.S.C. §991(b)(1)(B).

⁸³Marc L. Miller. & Ronald F. Wright, *Your Cheatin’ Heart(land): The Long Search for Administrative Sentencing Justice*, 2 BUFF. CRIM. L. REV. 723 (1999).

⁸⁴See Stith & Koh, *supra* note X, at 284-85 (“It should come as a surprise to no one that in those areas where the statute is ambiguous or otherwise deliberately leaves important policy issues to the Commission, the Commission has generally chosen to increase the rigidity and complexity of its guidelines. It is no accident that judges have found it difficult to depart from the guidelines; this is precisely what Congress intended.”).

⁸⁵In addition to creating the Guidelines, the SRA also abolished federal parole and instituted supervised release in its place. Supervised release is meant “to assist individuals in their transition to community life,” and “fulfills rehabilitative ends, distinct from those served by incarceration.” *United States v. Johnson*, 529 U.S. 53, 59 (2000). The term of supervised release is imposed along with a prison term at the time of sentencing. Under USSG §5D1.1(a), a sentencing court “shall order a term of supervised release to follow imprisonment when a sentence of imprisonment of more than one year is imposed” United States Sentencing Commission, Guidelines Manual, §5D1.1(a). Subsequent provisions provide for the minimum term of supervised release for defendants convicted of different classes of felonies with no statutory minimum. *Id.* §5D1.2(a). Certain types of felony offenses, such as drug trafficking offenses and sex offenses, are associated with mandatory terms of supervised release. See, e.g., 21 U.S.C. §841, 846, 960 and 963 for mandatory terms of supervised release for drug trafficking, 18 U.S.C. §3583(k) for mandatory terms of supervised release for sex offenses involving minors, and 18 U.S.C. §3583(a) for mandatory terms of supervised release for domestic violence offenses. During supervised release, defendants must regularly report to their probation officer, submit to drug testing, and remain within specified areas. If a defendant violates the terms of his/her supervised release, he/she faces revocation of supervised release which can lead to additional imprisonment. See United States Sentencing Commission, Guidelines Manual, §7B1.4.

⁸⁶H.R. REP. NO. 1017, 98th Congress, 2nd Session, at 93-94 (1984).

⁸⁷See Miller & Wright, *supra* note X, at 730 (The Commission allowed judges to depart from the guideline when the case fell within the “heartland” but the concept was left highly general.)

⁸⁸518 U.S. 81, 99 (1996). The Court in *Koon* stated that Congress “did not intend, by establishing limited appellate review, to vest in appellate courts wide-ranging authority over district court sentencing decisions.” *Id.* at 97.

⁸⁹United States Sentencing Commission, Guidelines Manual, Chapter Two.

⁹⁰*Id.* §2B2.3.

⁹¹*Id.* §2A4.1.

and adjustments. Relevant adjustments include the amount of loss involved in the offense, use of a firearm, the age or condition of the victim, etc.⁹² Chapter Three of the Guidelines allows for further adjustments based on aggravating or mitigating factors, such as a defendant's acceptance of responsibility.⁹³

The criminal history category reflects the frequency and severity of a defendant's prior criminal convictions. To determine a defendant's criminal history category, a judge adds points for prior sentences in the federal system, fifty state systems, all territories and foreign or military courts.⁹⁴ For example, three points are added for each prior sentence of imprisonment exceeding one year and one month, and two points are added for each prior sentence of imprisonment of at least 60 days and less than one year and one month.⁹⁵ Two points are also added if the defendant committed the instant offense under any criminal justice sentence.⁹⁶

The intersection of the final offense level and criminal history category yields a fairly narrow Guidelines recommended sentencing range, where the top of the range exceeds the bottom by the greater of either 6 months or 25%. If a judge determines that there are aggravating or mitigating circumstances that warrant a departure from the Guidelines, he/she would have to justify his/her reasons for departure to the appellate court,⁹⁷ but in general the Guidelines were treated as sufficiently mandatory prior to *Booker*.⁹⁸ After the imposition of a sentence, the government is permitted to appeal any sentence resulting in a departure below the Guidelines range, and the defendant can appeal an upward departure.⁹⁹

There are numerous other ways in which Congress has attempted to limit unwarranted disparities in sentencing. Beginning in 1984, and subsequently 1986 and 1988, Congress enacted a series of mandatory minimum statutes directed at drug and firearms offenses.¹⁰⁰ Mandatory minimums were also applied to recidivist offenders, through the Armed Career Criminal Act,¹⁰¹ enhancements for career offenders,¹⁰² and enhancements for repeat and dangerous sex offenders.¹⁰³

⁹² *Id.*

⁹³ For instance, the Guidelines allow for a decrease in base offense level for a defendant's acceptance of responsibility under §3E1.1 or for minimal participation in the offense under §3B1.2. Base offense level is increased for defendants who obstruct or impede the administration of justice under §3C1.1.

⁹⁴ United States Sentencing Commission, Guidelines Manual, §4.1 ("A defendant with a record of prior criminal behavior is more culpable than a first offender and thus deserving of greater punishment. Greater deterrence of criminal conduct dictates that a clear message be sent to society that repeated criminal behavior will aggravate the need for punishment with each recurrence. To protect the public from further crimes of the particular defendant, the likelihood of recidivism and the future criminal behavior must be considered. Repeated criminal behavior is an indicator of a limited likelihood of successful rehabilitation.").

⁹⁵ *Id.* §4.1.

⁹⁶ *Id.* §4.1.

⁹⁷ 18 U.S.C. §3553(b) ("the court shall impose a sentence of the kind, and within the range, referred to in subsection (a)(4) unless the court finds that there exists an aggravating or mitigating circumstance of a kind, or to a degree, not adequately taken into consideration by the Sentencing Commission in formulating the guidelines that should result in a sentence different from that described").

⁹⁸ The Court in *Booker* noted that "[t]he Guidelines as written ... are not advisory; they are mandatory and binding on all judges" and therefore "have the force and effect of laws." *Booker*, 543 U.S. at 233.

⁹⁹ 18 U.S.C. §3742 (a)-(b).

¹⁰⁰ Anti-Drug Abuse Act of 1986 (Pub. L. 99-570, 100 Stat. 3207); Pub. L. No. 98-473, §1005(a), 98 Stat. 1837, 2138-39 (1986).

¹⁰¹ Guidelines Manual §4B1.2; 18 U.S.C. §924(e). The Armed Career Criminal Act (ACCA) imposes a minimum 15-year term of imprisonment for defendants convicted of unlawful possession of a firearm under 18 U.S.C. §922(g), with three prior state or federal convictions for violent felonies or serious drug offenses.

¹⁰² §4B1.1; 28 U.S.C. §944(h) (mandating that the Commission impose imprisonment "at or near the maximum term authorized" for defined "career" offenders).

¹⁰³ §4B1.5; 18 U.S.C. §§2247, 2426.

In 2003, Congress passed the PROTECT Act to curtail judicial departures due to a concern that the standard for appellate review of departures had led to undesirably high rates of downward departures for child sex offenses.¹⁰⁴ Under the Feeney Amendment to the PROTECT Act, judicial departures were only allowed for certain reasons outlined in the Guidelines Manual.¹⁰⁵ Additionally, the Feeney Amendment to the PROTECT Act replaced the prior abuse of discretion standard of review for downward departures with *de novo* review by overturning the Supreme Court's holding in *Koon*.

3.2.2. Challenges to the Mandatory Guidelines Regime

The initial challenge to the federal sentencing began with the “watershed” ruling in *Apprendi v. New Jersey*.¹⁰⁶ In *Apprendi*, the Supreme Court found a New Jersey hate crime statute unconstitutional because it authorized judges to impose higher sentences based on facts other than those submitted to a jury, and proved beyond a reasonable doubt.¹⁰⁷

These principles were subsequently applied to the constitutionality of mandatory sentencing guidelines, first questioned in reference to the Washington State Sentencing Guidelines. In *Blakely v. Washington*, the Supreme Court held that the Sixth Amendment right to a jury trial prohibited judges from increasing a defendant's sentence beyond the statutory maximum based on facts other than those decided by the jury beyond a reasonable doubt.¹⁰⁸ As a result, Washington's mandatory sentencing guidelines were struck down. While the majority opinion in *Blakely* emphasized that the decision was not passing judgment on the constitutionality of the Federal Sentencing Guidelines,¹⁰⁹ the parallels were apparent and shortly after, the reasoning of *Blakely* was applied to the Guidelines.

In *United States v. Booker*, the mandatory Federal Sentencing Guidelines were also found to be unconstitutional under the Sixth Amendment by mandating judicial fact-finding for determining sentencing ranges.¹¹⁰ The *Booker* ruling, however, did not apply to Congressionally-enacted mandatory minimum sentences.¹¹¹ Rather than invalidate the Guidelines wholly, or prescribe an enhanced role for jury fact-finding, the Court held in a separate remedial decision led by Justice Breyer, that the remedy for the Sixth Amendment violation was to declare the Guidelines no longer mandatory but “effectively advisory.”¹¹² The Court explained that “district courts, while not bound to apply the Guidelines, must consult those Guidelines and take them into account when sentencing.”¹¹³

¹⁰⁴Pub. L. 108-21, 117 Stat. 650, S. 151, 2003.

¹⁰⁵For certain offenses, such as child abduction and child sex offenses, the PROTECT Act amended 18 U.S.C. §3553(b) to only allow the sentencing court to depart downwards if there are mitigating circumstances of a kind or to agree that has been “affirmatively and specifically identified” as permissible grounds for downward departure. The PROTECT Act also amended the Guidelines Manual §5K2.0 to state that the “the grounds enumerated in this Part K of Chapter Five are the sole grounds that have been affirmatively and specifically identified as a permissible ground of departure.”

¹⁰⁶530 U.S. 466, 425 (2000) (O'Connor, J., dissenting) (calling the *Apprendi* decision a “watershed change in constitutional law”).

¹⁰⁷*Id.* at 468-69, 490.

¹⁰⁸542 U.S. 296 (2004).

¹⁰⁹*Id.* at 305, n.9 (“The Federal Guidelines are not before us, and we express no opinion on them”).

¹¹⁰543 U.S. 220, 226-27, 243-44 (2005) (Stevens, J., writing for the Court).

¹¹¹CITE.

¹¹²*Booker*, 543 U.S. at 245 (Breyer, J., writing for the Court). Similarly, the provisions on supervised release also became advisory, although the USSC states that the majority of courts continue to impose at least the minimum terms in §5D1.2.

¹¹³*Booker*, 543 U.S. at 264.

In the immediate aftermath of *Booker*, district courts took varied approaches in applying *Booker*.¹¹⁴ Some courts sentenced with minimal consideration of the applicable Guidelines range, while others treated the Guidelines as a dominant factor.¹¹⁵ Circuit courts later reached a consensus that sentencing must begin with the calculation of the applicable Guidelines range.¹¹⁶ Today, after a sentencing judge has calculated the applicable Guidelines range, he or she must consider 7 factors before imposition of punishment: (1) the nature and circumstances of the offense and the history and characteristics of the defendant, (2) the need for the sentence imposed, (3) the kinds of sentences available, (4) the kinds of sentence and the sentencing range established, (5) any pertinent policy statement issued by the Sentencing Commission, (6) the need to avoid unwarranted sentence disparities among defendants with similar records who have been found guilty of similar conduct, and (7) the need to provide restitution to any victims of the offense.¹¹⁷ After consideration of all the factors, the sentencing judge is instructed to “impose a sentence sufficient, but not greater than necessary, to comply with” the basic goals of sentencing.¹¹⁸

Subsequent Supreme Court decisions furthered weakened the effect of the Guidelines on criminal sentencing. In *Rita v. United States*, the Court held that a sentence within the Guidelines recommended range could be presumed “reasonable” because a “judge who imposes a sentence within the range recommended by the Guidelines thus makes a decision that is fully consistent with the Commission’s judgment in general.”¹¹⁹ In *Gall v. United States*, the Court further held that federal appeals courts could not presume that a sentence outside the range recommended by the Guidelines was unreasonable, reducing the degree of appellate review.¹²⁰ The Court in *Gall* concluded that in reviewing a sentence outside the Guidelines range, an appellate court could consider the extent of deviation from the Guidelines, but must give “due deference to the district court’s decision that the §3553(a) factors, on a whole, justify the extent of the variance.”¹²¹ In the aftermath of *Gall*, appellate courts could only review sentencing decisions under the more deferential abuse of discretion standard. Concurrent with the *Gall* decision, the Supreme Court confirmed the holding in *Booker* as applied to cases involving possession, distribution and manufacture of crack cocaine.¹²² The Court in *Kimbrough* held that federal district court judges have the discretion to impose sentences outside the

¹¹⁴ See Gilles R. Bissonnette, “Consulting” the Federal Sentencing Guidelines After *Booker*, 53 UCLA L. REV. 1497, 1521-22 (2006) (arguing that *Booker* left open how much sentencing judges could deviate from the Guidelines).

¹¹⁵ See *id.* at 1522-32 (claiming that district courts have taken two approaches in applying *Booker*, the “substantial-weight” approach and the “consultative” approach). For an example of the two type of approaches taken by district courts, see, e.g. *United States v. Ranum*, 353 F. Supp. 2d 984, 987 (E.D. Wis. 2005) (noting that “*Booker* is not ... an invitation to do business as usual,” but that courts should consider all the factors in §3553(a)); cf. *United States v. Wilson*, 350 F. Supp. 2d 910, 925 (D. Utah 2005) (suggesting courts give “heavy weight” to Guidelines after *Booker*).

¹¹⁶ See, e.g., *United States v. Howard*, 454 F.3d 700, 703 (7th Cir. 2006) (court noting that after *Booker*, “the district court must first calculate the proper Guidelines range and then, by reference to the factors specified in 18 U.S.C. §3553(a), select an appropriate sentence”); *United States v. Crosby*, 397 F.3d 103, 113-114 (2d Cir. 2005) (“consideration of the Guidelines will normally require determination of the applicable Guidelines range, or at least identification of the arguably applicable ranges ... it would be a mistake to think that, after *Booker/Fanfan*, district judges may return to the sentencing regime that existed before 1987 and exercise unfettered discretion to select any sentence within the applicable statutory maximum and minimum”).

¹¹⁷ 18 U.S.C. §3553(a).

¹¹⁸ *Id.*

¹¹⁹ 551 U.S. 338, 347-50 (2007) (holding that “a court of appeals may apply a presumption of reasonableness to a district court sentence that reflects a proper application of the Sentencing Guidelines”).

¹²⁰ 552 U.S. 38 (2007).

¹²¹ *Id.* at 51 (arguing the an appellate court’s disagreement with the appropriateness of a sentence is “insufficient to justify reversal”).

¹²² *Kimbrough v. United States*, 552 U.S. 85 (2007).

recommended Guidelines range due to policy disagreements with the Sentencing Commission.¹²³

3.3. Framework, Data, and Methods

3.3.1. Judicial Behavior in Criminal Sentencing

A prominent consideration underlying judicial behavior is an obligation to “follow the law.”¹²⁴ However, judges have additional motivations that affect their decision-making.¹²⁵ Scholars have suggested that judges care about variety of factors such as public recognition, leisure, and reputation.¹²⁶ In addition, judges have preferences for sentencing according to their personal tastes, but may face costs associated with exercising discretion.¹²⁷ In the context of criminal sentencing, a judge may prefer to sentence a defendant based on personal, political or ideological goals, rather than the mandated Guidelines sentence.¹²⁸

How does the sentencing regime affect judicial behavior? Given individual judge-specific preferences for sentencing, if judges were left unconstrained in the exercise of discretion, one would likely observe large inter-judge disparities. Consistent with this prediction, scholars have suggested that the large variances in federal sentences prior to the adoption of the Guidelines were likely due to differing judicial attitudes regarding rehabilitation and deterrence.¹²⁹ At the other end of the spectrum, judges who are restricted in the exercise of discretion would be unable to sentence fully according to their preferences. Thus, the adoption of determinate sentencing under the Guidelines introduced a mechanism by which to constrain judges, likely explaining studies finding reduced inter-judge sentencing disparity after the promulgation of the Guidelines.¹³⁰

In addition to a restraint on free sentencing imposed by the sentencing regime, another constraint comes from the prospect of appellate review. Judges who depart from the Guidelines incur economic and social costs from deviating.

¹²³*Id.* (granting sentencing judges explicit permission to deviate from the Guidelines based on disagreement with the disparate treatment of crack and powder cocaine offenses - the so called “100-to-1 ratio”).

¹²⁴*See, e.g.,* Lewis A. Kornhauser, *Judicial Organization and Administration*, 7 *Encyclopedia of Law and Economics* 27 (2000).

¹²⁵The economic analysis of judicial behavior builds on work by Judge Richard A. Posner. *See, e.g.* Hon. Richard A. Posner, *Judicial Behavior and Performance: An Economic Approach*, 32 *FLA. ST. U. L. REV.* 1259 (2005); *see also* Hon. Richard A. Posner, *HOW JUDGES THINK* (2008); Lee Epstein, William M. Landes & Richard A. Posner, *THE BEHAVIOR OF FEDERAL JUDGES: A THEORETICAL AND EMPIRICAL STUDY OF RATIONAL CHOICE* (2013). *See* Nicola Gennaioli & Andrei Shleifer, *Judicial Fact Discretion*, 37 *J. LEG. STUD.* 1 (2008) for a theoretical economic model of judicial discretion in fact determination.

¹²⁶Federal district judges occupy a unique position because most district judge appointments are terminal, thus “insulat[ing] the judges from the normal incentives and constraints that determine the behavior of rational actors, except for the relative handful of judges who are ambitious for promotion to the court of appeals.” Posner, *Judicial Behavior and Performance: An Economic Approach*, *supra* note X, at 1260, 1269 (noting that a judge likely care more about leisure and public recognition relative to income, compared to average practicing attorneys).

¹²⁷Posner, *Judicial Behavior and Performance: An Economic Approach*, *supra* note X, at 1269-70 (“deciding a particular case in a particular way might increase the judge’s utility just by the satisfaction that doing a good job produces ... [or] by advancing a political or ideological goal”).

¹²⁸Indeed, federal district court judges have expressed a great degree of dissatisfaction with the Guidelines. In a 2010 survey of federal district judges, 65% of judges indicated that they thought the departure policy statements in the Guidelines Manual were too restrictive, suggesting that many judges would prefer to deviate from the Guidelines. U.S. Sentencing Commission, *RESULTS OF SURVEY OF UNITED STATES DISTRICT JUDGES JANUARY 2010 THROUGH MARCH 2010* (June 2010), (Question 14, Table 14).

¹²⁹*See* Posner, *Judicial Behavior and Performance: An Economic Approach*, *supra* note X, at 1270 (inferring from the “extraordinary variance” in federal sentences prior to the promulgation of the Guidelines that differences in sentence lengths were due to judicial attitudes on responsibility and deterrence); *see also* Brian Forst & Charles Wellford, *Punishment and Sentencing: Developing Sentencing Guidelines Empirically From Principles of Punishment*, 33 *RUTGERS L. REV.* 799, 801-804 (1981) (documenting disagreement between judges regarding five goals of sentencing: general deterrence, special deterrence, incapacitation, rehabilitation, and just deserts).

¹³⁰*See supra* notes 7-10.

A high reversal rate is not only administratively burdensome, but also potentially harms a trial judge's prospects for promotion to the appeals court.¹³¹ Indeed, under the mandatory sentencing regime, departures from the Guidelines were relatively rare.¹³² After the Feeney Amendment of PROTECT Act, which subjected district court judges to *de novo* review for departures from the Guidelines, suggesting that inter-judge sentencing disparity during this period may have been particularly reduced.¹³³

Given the countervailing forces of judge sentencing preferences against costs of exercising discretion, what is the theoretical prediction of the impact of *Booker*, *Rita*, *Gall*, and *Kimbrough* on inter-judge disparities? Following *Booker*, the total cost of exercising discretion fell substantially for judges who wanted to depart from the Guidelines sentence as the Guidelines were rendered advisory. This major shift in sentencing may be accompanied with increases in inter-judge disparity. However, to the extent the relative cost associated with *de novo* appellate review was still binding, judges may have been hesitant to alter their practices. Indeed, not until *Rita*, *Gall*, and *Kimbrough* did the Court reduce the level of appellate review from *de novo* to substantial abuse of discretion, intuitively lowering the probability of appellate reversal.¹³⁴ Thus, one might expect larger increases in inter-judge disparities following *Rita*, *Gall*, and *Kimbrough*. Nevertheless, given the *Rita* presumption of reasonableness attached to within range sentences, the Guidelines provide a safe harbor from appellate scrutiny.

There are other reasons why judicial behavior and inter-judge disparities may not change much after *Booker* and its progeny. First, judges may become acculturated to the Guidelines if they have had substantial experience sentencing under the previous Guidelines regime.¹³⁵ Acculturation would suggest that judges with greater exposure to Guidelines sentencing would be less likely to depart from the Guidelines in the aftermath of *Booker*.

Another potential mechanism is anchoring, a type of cognitive bias in which decision-makers rely heavily on one piece of information and fail to make rational adjustments.¹³⁶ Judge Nancy Gertner of the District of Massachusetts

¹³¹ See Posner, *Judicial Behavior and Performance: An Economic Approach*, *supra* note X, at 1270-71; see also Stephen J. Choi, Mitu Gulati, & Eric A. Posner, *What do Federal District Judges Want?: An Analysis of Publications, Citations, and Reversals*, University of Chicago John M. Olin Law & Economics Working Paper Series Paper No. 508, at 3-4 (2009) (judges care about minimizing workload and maximizing reputation by avoiding appellate reversal); Evan H. Caminker, *Precedent and Prediction: The Forward-Looking Aspects of Inferior Court Decisionmaking*, 73 TEX. L. REV. 1, 77-78 (1994) (describing anecdotal evidence that lower court judges dislike being reversed on appeal due to professional reputation, advancement, judicial power); Richard S. Higgins & Paul H. Rubin, *Judicial Discretion*, 9 J. LEGAL STUD. 129, 130 (1980).

¹³² The rate of departure from the Guidelines was less than 15% in the early 1990s.

¹³³ Recall that the Commission found that demographic differences under the mandatory guidelines regime were lower during the PROTECT Act. See *supra* note X [United States Sentencing Commission, DEMOGRAPHIC DIFFERENCES IN FEDERAL SENTENCING PRACTICES: AN UPDATE OF THE BOOKER REPORT'S MULTIVARIATE REGRESSION ANALYSIS (2010)].

¹³⁴ The probability of reversal on sentencing matters fell from 36% in 2006 (under *de novo* review), to 26% in 2008 (under abuse of discretion review). I calculate rate of appellate reversals using yearly data on the universe of criminal appeals from the USSC. Reversal is defined as all reversals and remands on appeals arising out of *Booker* sentencing issues.

¹³⁵ See Judge Nancy Gertner, *Supporting Advisory Guidelines*, 3 HARV. L. POL'Y REV. 261, 262 (2009) ("[A]fter twenty years of strict enforcement, the Federal Sentencing Guidelines have a gravitational pull on sentencing and are likely to shape the way judges view sentencing, even if they are now only advisory. Indeed, the greatest danger is not that judges will exercise their new discretion, but that they will not."); *Stith*, *The Arc of the Pendulum*, *supra* note X, at 1496-97 ("incumbent sentencing decision makers may be reluctant to regard as unreasonable the sentences they were obliged to seek and impose for two decades"); Frank O. Bowman, *Beyond Band-Aids: A Proposal for Reconfiguring Federal Sentencing After Booker*, 2005 CHI. LEGAL F. 149, X (2005) (arguing that advisory guidelines might still constrain judicial discretion "if for no other reason than that the federal bench has become acculturated to the Guidelines over the last seventeen years").

¹³⁶ See, e.g., Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124, 1130-31 (1974); see also Birte Englich et al., *Playing Dice with Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making*, 32 PERSONALITY & SOC. PSYCHOL. BULL. 188, 188 (2006) (experimental results showing that criminal sentences were higher if participants were confronted with a randomly high rather than a low anchor).

predicted that the Guidelines would still play a predominant role for all judges post *Booker* because “appellate courts have insisted that district court judges begin with - effectively, ‘anchor’ their decisions - in the Guidelines before considering anything else.”¹³⁷ Thus, to the extent that federal district judges are effectively anchored to the Guidelines, one may not observe much deviation in sentencing practices after *Booker*. Indeed, because district courts continued to calculate the applicable Guidelines range in the aftermath of *Booker*, scholars commented in the year following *Booker* that the federal sentencing system remained virtually unchanged.¹³⁸

3.3.2. Sentencing Data

This Article provides the first comprehensive empirical evidence on the impact of *Booker* and its progeny on inter-judge sentencing disparities. As noted previously, the Scott study is the only empirical study thus far, but is limited to the 2,262 cases sentenced by judges who served continuously from 2001 to 2008 in the Boston courthouse of the District of Massachusetts.¹³⁹ While the study is a first step in characterizing the extent to which inter-judge sentencing practices have changed in the aftermath of *Booker*, the Boston courthouse is likely unrepresentative of sentencing patterns in other courthouses across the United States. The presence of growing inter-judge sentencing disparities after *Booker* in the Boston courthouse may be the result of the particular caseload and judicial composition of that court, making conclusions that *Booker* has increased inter-judge sentencing disparities likely not generalizable across other courts.

Prior empirical research on inter-judge disparity and the impact of judicial demographics on sentencing practices has been hampered by the lack of judge identifiers.¹⁴⁰ Because cases are generally randomly assigned to judges within a district courthouse, judge identifiers allow one to compare judges within the same court and in the same time period, capturing judge differences in sentencing rather than different caseloads.¹⁴¹ However, the Sentencing Commission data does not identify the sentencing judge,¹⁴² In response, most researchers have resorted to using aggregate district-level variation in judicial demographics to control for judge sentencing preferences,¹⁴³ but this methodology could be flawed if districts with different judicial compositions differ in ways in that affect all judges within the district.¹⁴⁴ A

¹³⁷See Judge Nancy Gertner, *What Yogi Berra Teaches About Post-Booker Sentencing*, 115 YALE L.J. POCKET PART (2006) 137, 138-40 (describing the Guideline Manual as a “ready-made anchor”).

¹³⁸See Douglas A. Berman, *Tweaking Booker: Advisory Guidelines in the Federal System*, 43 HOUSTON L. REV. 341, at 349 (2006) (“a culture of guideline compliance has persisted after *Booker*”). Berman also suggests that Commission data in the year after *Booker* indicates that “federal sentencing judges are exercising their new discretion relatively sparingly.” *Id.* at 351.

¹³⁹Scott, *supra* note X, at 17.

¹⁴⁰The Anderson et al. study is the only empirical work with comprehensive sentencing data with judge identifiers in the past 25 years. See Anderson et. al, *supra* note X, at 287.

¹⁴¹According to the Administrative Office of the United States Courts, “The majority of courts use some variation of a random drawing.”

¹⁴²United States Sentencing Commission, GUIDE TO PUBLICATIONS & RESOURCES 2007-2008 45 (2007), available at <http://www.ussc.gov/publicat/Cat2005.pdf> (“Pursuant to the policy on public access to Sentencing Commission documents and data, all case and defendant identifiers have been removed from the data.” (internal citation omitted)).

¹⁴³Fischman & Schanzenbach, *Strategic Judging*, *supra* note X, at X (relying only on district-level variation in observable characteristics of judges).

¹⁴⁴For instance, a district with a greater percentage of Democrat judges could be different from other districts. It may be that both Democrat and Republican judges within the district sentence differently from judges in other districts, so any effect cannot be solely attributable to Democrat judges.

few researchers have resorted to hand matching sentencing data from the Sentencing Commission with Public Access to Court Electronic Records (PACER), but due to the lengthy matching process, sample sizes have been severely limited.¹⁴⁵

This paper is the first in over 25 years to match sentencing data from fiscal years 2000-2009 to judge identifiers in all 94 district courts, allowing for a comprehensive look at inter-judge sentencing disparities after *Booker*. In order to overcome the lack of judge identifiers in sentencing data, I utilize datasets from three sources: The United States Sentencing Commission, the Transactional Records Access Clearinghouse, and the Federal Judicial Center. I describe each dataset in turn.

United States Sentencing Commission - I use data from the United States Sentencing Commission (USSC) on records of all federal offenders sentenced pursuant to the Sentencing Guidelines and Policy Statements of the SRA of 1984 in fiscal years 2000-2009 (October 1, 1993 - September 30, 2009). The USSC provides detailed sentencing data on federal defendants, but defendant and judge identifiers are redacted.¹⁴⁶ The dataset contains information from numerous documents on every offender: Indictment, Presentence Report, Report on the Sentencing Hearing, Written Plea Agreement (if applicable), and Judgment of Conviction.

Court characteristics include the district court and circuit in which sentencing occurred, in addition to the sentencing month and year.¹⁴⁷ Demographic variables include defendant's race, gender, age, citizenship status, educational attainment, and number of dependents. The primary offense variable is the primary offense type for the case generated from the count of conviction with the highest statutory maximum.¹⁴⁸ Data is also available on whether the offense carries a mandatory minimum sentence under various statutes, and whether departures from the statutory minimum are granted, either under a substantial assistance motion or application of the safety valve. Offense level variables include the base offense level, the base offense level after Chapter Two adjustments and the final offense level after Chapter Three adjustments. Criminal history variables include whether the defendant has a prior criminal record (first time offender or prior offender), and whether armed career criminal status, career offender status, or repeat and dangerous sex offender status is applied.¹⁴⁹

For each offender, there is a computed Guidelines range, and a Guidelines range adjusted for applicable mandatory minimums. Sentencing outcomes include imprisonment or probation, sentence length, and length of supervised release. From these variables, I construct indicator variables for above range and below range departures from the

¹⁴⁵See, e.g., Schanzenbach & Tiller, *Reviewing the Sentencing Guidelines: Judicial Politics, Empirical Evidence, and Reform*, 75 U. CHI. L. REV. 715 (2008); Scott, *supra* note X, at 15-16 (describing the PACER method used to match records to sentencing data).

¹⁴⁶Over 90% of felony defendants in the federal criminal justice system are sentenced pursuant to the SRA of 1984 and all cases are assessed to be constitutional.

¹⁴⁷USSC data prior to 2004 actually includes information on the exact sentencing day, but this variable is not available in later years. Information is also collected on the Guidelines amendment year used in calculations. All results are robust to controlling for amendment year, although the results presented in this paper do not include this control.

¹⁴⁸There are a total of 35 offense categories in the dataset. The most common offense is drug trafficking, followed by immigration, fraud, firearms, and larceny.

¹⁴⁹Data is also collected on the total number of criminal history points applied and the final criminal history category.

Guidelines.¹⁵⁰

Transactional Records Access Clearinghouse - The Transactional Records Access Clearinghouse (TRAC) provides less comprehensive sentencing records obtained from detailed federal records and information from the Justice Department and the Office of Personnel Management. Defendant, offense characteristics and Guidelines application information is not included, but defendants are linked with sentencing judge.¹⁵¹

Federal Judicial Center - Finally, I obtain demographic information on federal district court judges. Federal district judges are Article III judges who serve life term tenures. New appointments are generally made when a judge retires or dies.¹⁵² As of the current day, there are a total of 677 authorized Article III district judgeships.¹⁵³ The number of federal district judgeships in each district court varies. The largest district court is the Southern District of New York with 28 authorized judgeships. The majority of other district courts have between 2-7 district court judgeships. I collect information on judge race, gender, affiliation of appointing president, tenure, whether the judge was appointed prior to the adoption of the Guidelines in 1987, and whether the judge was appointed after *Booker*.

3.3.3. Matching

First, I match sentencing records from the USSC to TRAC data. By district court, matching is conducted on several key variables that can uniquely identify each record: sentencing year, sentencing month, offense type,¹⁵⁴ sentence length in months, probation length in months, amount of monetary fine, whether the case ended by trial or plea agreement, and whether the case resulted in a life sentence.¹⁵⁵

Finally, I match the USSC and TRAC combined data to judge biographical data from the Federal Judicial Center. I successfully match over 90% of all USSC cases from fiscal years 2000-2009, resulting in a matched dataset of 636,063 cases representing 91 district courts (hereinafter “full sample”).¹⁵⁶

3.3.4. Testing for Random Assignment

In an ideal experiment to test the impact of *Booker*, one would randomly allocate the treatment - sentencing under an advisory Guidelines - to certain groups of judges. In this hypothetical, a group of judges within each district court would be randomly selected into the treatment group, while the others would comprise the control group. Because

¹⁵⁰An above range departure is defined as 1 for a defendant who received a sentence above the maximum Guidelines recommended range. Similarly, a below range departure is defined as 1 for a defendant who received a sentence below the minimum Guidelines recommended range.

¹⁵¹TRAC has compiled records on the criminal cases and the civil matters handled by federal district court judges in each of the 94 federal judicial districts through over 20 years of FOIA requests.

¹⁵²On a few occasions, Congress has also increased the number of judgeships within a district in response to changing population or caseload, the last additions for certain district courts taking place in 2002.

¹⁵³67 positions are currently vacant. Authorized judgeships only refer to full-time non-senior status judges.

¹⁵⁴Data from USSC are coded to correspond with the offense categories in the TRAC data.

¹⁵⁵These match variables are to those used by previous scholars under the PACER method. See Schanzenbach & Tiller, *supra* note 72, 729 (matching USSC data with PACER records on date and length of the sentence, and when necessary, the amount of any fine, the offense type, and the Hispanic ethnicity of the defendant).

¹⁵⁶The Federal Judicial Center does not collect demographic information on judges in 3 districts: Guam, Virgin Islands, and Northern Mariana Islands.

of the random assignment of the “*Booker*” treatment, any differences in caseload composition or judge characteristics would be on average the same across both treatment and control groups. As a result, a straightforward comparison of the sentencing practices between judges in the treatment group (those who sentence under an advisory Guidelines), and the judges in the control group (those who sentence under a mandatory Guidelines), would capture the causal effect of greater judicial discretion via *Booker* on inter-judge differences in sentencing.

Unfortunately, this hypothetical experiment does not exist because the Supreme Court’s decision in *Booker* applied to all judges. However, one can utilize the quasi-experiment created by the timing of the *Booker* decision. Assuming that judges within the same district courthouse are randomly assigned cases from the same underlying caseload, one can compare inter-judge disparities before *Booker* to inter-judge disparities after *Booker*. If there are no other contemporaneous changes that affect inter-judge sentencing, an increase in inter-judge disparities is attributable to changes in judicial behavior, rather than underlying differences in case composition. Moreover, random assignment of cases can also lead to estimates of the relationships between judicial demographics such as gender and experience on sentencing practices.

The crucial assumption underlying the validity of the quasi-experiment is the random assignment of cases to judges.¹⁵⁷ According to the Administrative Office of the United States Courts, “[t]he majority of courts use some variation of a random drawing” as prescribed by local court orders and while courts use different procedures for assigning cases to judges, “[m]ost district and bankruptcy courts use random assignment, which helps to ensure a fair distribution of cases and also prevents ‘judge shopping,’ or parties’ attempts to have their cases heard by the judge who they believe will act most favorably.”¹⁵⁸ Each district court is authorized to specify its own methods of case assignment.¹⁵⁹

However, random assignment may be violated in some instances. For example, senior status judges with reduced caseloads may select the type of cases they hear during the year,¹⁶⁰ and some courts assign certain types of cases to particular judges.¹⁶¹ Moreover, even if a court has local rules and orders that specify the use of random assignment, empirically testing for random assignment is important because random assignment can be violated if individuals seek

¹⁵⁷ As mentioned previously, the Anderson et. al, Hofer, Scott, and TRAC studies all rely on random case assignment. See *supra* notes X.

¹⁵⁸ For a description of judge assignment methods, see the Federal Judicial Center FAQs, available at <http://www.fjc.gov/federal/courts.nsf/>.

¹⁵⁹ Under 28 U.S.C. §137, “[t]he business of a court having more than one judge shall be divided among the judges as provided by the rules and orders of the court. The chief judge of the district court shall be responsible for the observance of such rules and orders, and shall divide the business and assign the cases so far as such rules and orders do not otherwise prescribe.” For example, in the Arizona district court, “the Clerk must assign criminal cases to District Judges within each division by automated random selection and in a manner so that neither the Clerk nor any parties or their attorneys will be able to make a deliberate choice of a particular Judge.” AZ. L. R. Crim. 5.1(a). In the Northern District of California, “[c]ases shall be assigned blindly and at random by the clerk by means of a manual, automated or combination system approved by the judges of the court.” CA General Order No. 44. In Colorado, “[t]he clerk shall maintain a computerized program to assure random and public assignment of new cases on an equal basis among the judicial officers.” D.C. Colo. L Civ R 40.1.

¹⁶⁰ 28 U.S.C. §294 governs the assignment of cases to senior status judges. See, e.g., MA General Order 10-04 §4.2 (“A senior judge may limit the category of cases assigned to him or her or may select a special category of cases for assignment. For example, a senior judge may elect not to be assigned criminal cases or may elect to be assigned only patent cases.”)

¹⁶¹ For instance, the Southern District of New York assigns civil and criminal cases such that all judges, “except the chief judge, shall be assigned substantially an equal share of the categories of cases of the court over a period of time. There shall be assigned or transferred to the chief judge such matters as the chief judge is willing and able to undertake, consistent with the chief judge’s administrative duties.” Thus, assignment is based on equal caseload, rather than pure random assignment, and the chief judge is exempted from the rules. See Rules for the Division of Business Among District Judges, available at <http://www.nysd.uscourts.gov/courtrules.php>.

to game the system.¹⁶² For instance, some courts have decried situations in which high profile cases are given to judges viewed as favorably disposed to one side.¹⁶³

In order to dispose of potential violations to random case assignment, I empirically test for random assignment. To begin, I employ several sample restrictions. First, I drop judges who were formally retired prior to the beginning of the dataset in 2000 to remove the possibility of non random assignment to senior status judges who continued to hear cases during the sample period. Second, I drop judges and district courthouses with annual caseloads of less than 25 cases in order to obtain a sufficient number of cases per judge for statistical power.¹⁶⁴ Finally, I drop district courthouses with only one active judge.

With this restricted sample, I test for randomly assignment using the matched USSC, TRAC, and Federal Judicial Center data from 2000-2009.¹⁶⁵ If cases are randomly assigned to judges, then judges should see on average, cases with the same distribution of predetermined defendant characteristics. To test for random assignment, I regress the five defendant characteristics on district courthouse by sentencing year fixed effects, sentencing month fixed effects and judge fixed effects. The five defendant characteristics include: gender, age, a black race indicator, number of dependents, and an indicator for less than a high school degree. Intuitively, there should be no significant correlation between a particular judge and defendant characteristics if cases are randomly assigned.

However, in testing the random assignment of defendants across these five characteristics, I encounter the problem that defendant characteristics are not fully independent. For instance, black defendants are also likely to have completed less than a high school degree. To address the confounding nature of these characteristics, I use seemingly unrelated regression (SUR) to test for random assignment. SUR allows me to test random assignment simultaneously for all the five defendant characteristics, addressing correlations.¹⁶⁶

SUR can be formally described as the regression model:

$$Y_{dtmj} = \alpha_0 + \gamma_d + \delta_t + \gamma_d * \delta_t + \lambda_m + \kappa_j + \epsilon_{dtmj} \quad (3)$$

¹⁶² See J. Robert Brown Jr. & Allison Herren Lee, *Neutral Assignment of Judges at the Court of Appeals*, 78 Texas L. Rev 1037 (2000) (describing nonrandom assignment in federal courts of appeals).

¹⁶³ See *United States v. Pearson*, 203 F.3d 1243, 1264 (10th Cir. 2000) (“If a judge receives case assignments not through some neutral system, but rather because of prosecutors’ opinion that he or she is more favorably disposed to the government’s arguments than another judge in the same district, then a judge’s caseload might be based in part on prosecutors’ evaluations of judicial performance.”). For recent allegations of “gaming” the random assignment system, see William Safire, *Essay: Norma The Plumber*, N.Y. TIMES, July 31, 2000, available at <http://www.nytimes.com/2000/07/31/opinion/essay-norma-the-plumber.html> (allegations that Chief Federal Judge of the U.S. District Court (U.S.D.C.) went “off the wheel” to assign a politically sensitive case in a non-random fashion); see also Dan Fitzpatrick, *Bank of America Manages to Avoid Judge Rakoff*, WALL ST. J., May 17, 2010, available at <http://online.wsj.com/article/SB10001424052748703699804575247132437874588.html> (nonrandom assignment of Bank of America case in the S.D.N.Y.).

¹⁶⁴ Results are robust to choice of caseload minimums, but follow the convention in prior literature. The Scott study excludes judges who imposed less than 25 sentences in a two year period. See Scott, *supra* note X, at 17. Similarly, Anderson et. al exclude judges who sentence less than 30 cases during a two year period. Anderson et al., *supra* note X, at 287.

¹⁶⁵ See Appendix A for details.

¹⁶⁶ Testing each characteristic individually would result in incorrect standard errors if the demographic characteristics are correlated. For a discussion of the SUR technique, see David H. Autor & Susan N. Houseman, *Do Temporary-Help Jobs Improve Labor Market Outcomes for Low-Skilled Workers? Evidence from “Work First”*, 2 AEJ: Applied Economics 96, 106-107 (2010).

where Y_{dtmj} is a characteristic of defendant i , sentenced in district court d in year t and month m , by judge j . The specification includes district court fixed effects (γ_d), sentencing year fixed effects (δ_t), sentencing month fixed effects (λ_m), and sentencing judge fixed effects (κ_j) to accurately compare cases assigned to judges in the same courthouse, and in the same year and month.

To formally test for random assignment, I test the null hypothesis of no judge effects - κ - using an F-test. The p-value for this F-test tests whether the defendant characteristics do not differ significantly among the cases that are assigned to district court judges in the same district courthouse, sentencing year, and sentencing month. A large p-value would signify the acceptance of the null hypothesis, and lead to the conclusion that random assignment was present. I drop all courthouses with F-test p-values less than 0.05, but results are robust to other cutoffs. I drop all courthouses that violate random assignment, resulting in a subsample of 163 courthouses from 73 district courts representing about 50% of the cases from 2000-2009, for a total of 270,334 cases (hereinafter “random sample”).

3.3.5. Trends in Sentencing

I first present graphical evidence of trends in sentence lengths and rates of below range departures using this matched dataset over the time period 2000-2009. Graphical analyses confirm that *Booker* did significantly alter the sentencing practices of judges. Figure 3.1 indicates that overall sentence lengths increased slightly prior to *Booker* but began to decrease afterwards, particularly for cases in which a mandatory minimum was not charged.¹⁶⁷

¹⁶⁷Cases defined as excluding mandatory minimums are those in which a statutory mandatory minimum was not charged.

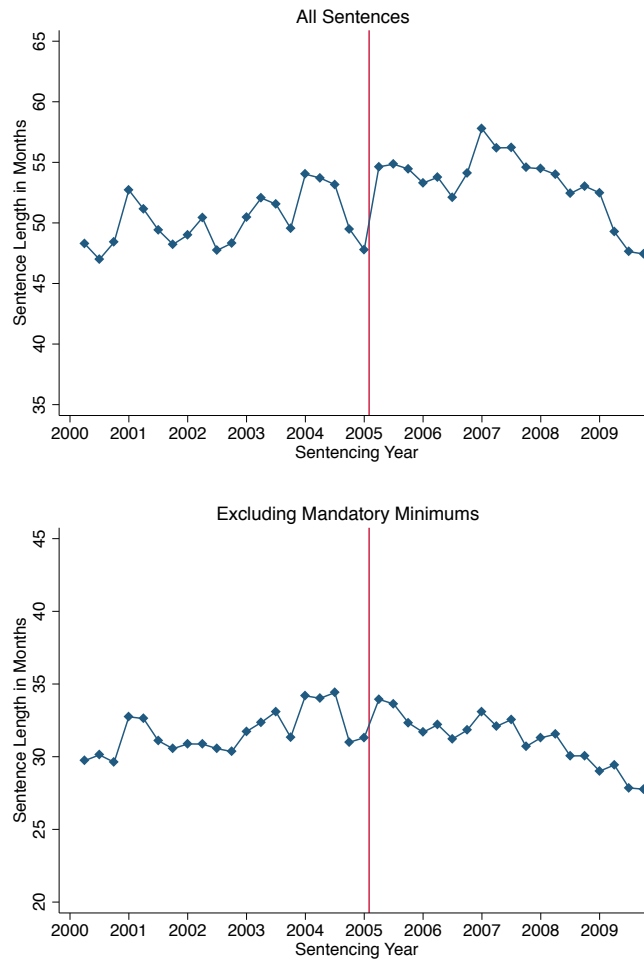


FIGURE 3.1. SENTENCE LENGTHS IN MONTHS

Notes: Data is from the random sample 2000-2009. Data points are quarterly averages.

Figure 3.2 reveals a trend of decreasing rates of below range departures prior to *Booker*, characterized by very low relative rates of departures in the PROTECT Act era. The decreasing trend in below range departures was significantly changed following *Booker*, which induced a sudden jump in the rate of departure, as well as increasing departure rates throughout *Booker*, *Rita*, *Gall*, and *Kimbrough*, back to pre-PROTECT era levels.

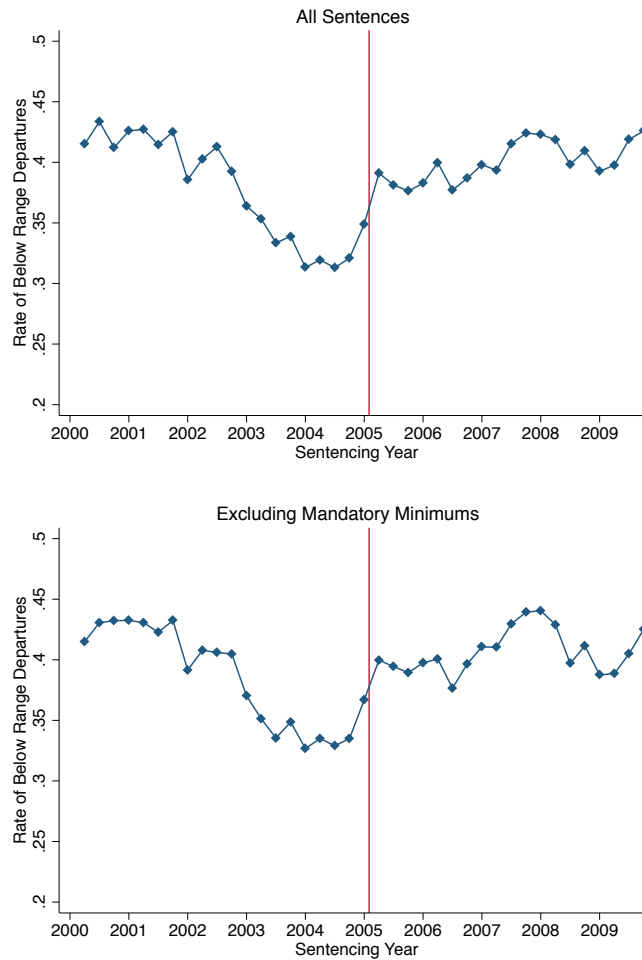


FIGURE 3.2. BELOW RANGE DEPARTURE RATES

Notes: Data is from the random sample 2000-2009. Data points are quarterly averages.

Figure 3.3 indicates a similar trend with respect to rates of non government sponsored below range departures, suggesting that judicial behavior has changed following the shift to an advisory guidelines regime.¹⁶⁸ However, while overall trends in sentencing have changed in the aftermath of *Booker*, and its progeny *Rita*, *Gall*, and *Kimbrough*, aggregate trends mask whether inter-judge variation has increased.

¹⁶⁸I define a nongovernment sponsored below range departure as a departure not resulting from a government substantial assistance motion.



FIGURE 3.3. NON GOVERNMENT SPONSORED BELOW RANGE DEPARTURES

Notes: Data is from the random sample 2000-2009. Data points are quarterly averages.

3.3.6. Measuring Inter-Judge Disparity - Analysis of Variance

I employ an analysis of variance methodology to identify changes in inter-judge disparity. Variants of this methodology has been used in the federal sentencing literature,¹⁶⁹ as well as in the economics literature on teacher value added.¹⁷⁰ The analysis of variance technique measures inter-judge dispersion in sentencing outcomes based on the variance of a judge-specific random variable.¹⁷¹

¹⁶⁹Studies using a similar methodology include Anderson et al., *supra* note 7, Joel Waldfogel, *Aggregate Inter-Judge Disparity in Sentencing: Evidence from Three Districts*, 4 FED. SENTENCING REP. 151 (1991); Abigail Payne, *Does Inter-judge Disparity Really Matter? An Analysis of the Effects of Sentencing Reforms in Three Federal District Courts*, 17 INT'L J. LAW & ECON. 337 (1997).

¹⁷⁰See, e.g. Raj Chetty et. al, *How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star*, 126 Q. J. ECON. 1593 (2011).

¹⁷¹This paper does not employ the statistical technique used by Scott. Scott regresses sentencing outcomes on dummy indicators for each sentencing judge, such that the corresponding R-squared measures the percentage of variance in the dependent variable that is explained by sentencing judge identity. Scott, *supra* note X, at 58. The author interprets an increase in the R-squared in time periods following *Booker* as indicative of growing inter-disparities. For instance, with regards to sentence lengths for cases excluding mandatory minimums, the author finds an increase in R-squared from 0.014 pre-*Booker* to 0.080 post *Kimbrough/Gall*. *Id.* at Table 2, at 34. However, the R-squared measure is problematic for two main reasons. First, the measure of R-squared does not have a straightforward interpretation in terms of actual inter-judge variation, in contrast

I implement an analysis of variance using a defendant-level random effects specification of the form:

$$Y_{ijk} = X_{ijk} * \beta + \alpha_k + v_{ijk}, \text{ where } v_{ijk} = \mu_{jk} + \epsilon_{ijk} \quad (4)$$

The dependent variable Y_{ijk} is the sentence length in months for defendant i assigned to judge j in district office k. The control variables X_{ijk} included defendant and crime characteristics and α_k are indicator variables for the district office in which sentencing occurred. The residual v_{ijk} is composed of a judge effect or value added that is constant for a judge over time, and an idiosyncratic defendant effect. I estimate the coefficients β and the judge effects μ by OLS. OLS estimation yields consistent estimate of β if the judge random effects are uncorrelated with the control variables X.

I estimate the magnitude of the judge effects under a mixed random effects specification, assuming that μ_{jk} is distributed $N(0, \sigma_\mu^2)$. Intuitively, within judge variance in v_{ijk} is used to estimate the defendant variance:

$$\hat{\sigma}_\epsilon^2 = \text{Var}(v_{ijk} - \bar{v}_{jk}) \quad (5)$$

The variance in the judge effect is the remainder:

$$\hat{\sigma}_\mu^2 = \text{Var}(v_{ijk}) - \hat{\sigma}_\epsilon^2 \quad (6)$$

The estimated standard deviation of judge effects on sentence length is $\sigma_\mu = X$, implying that a one standard deviation increase in judicial harshness raises a defendants sentence by X months. Because the regression specification includes district office fixed effects, this measure represents the impact of being assigned a judge one SD higher in harshness of the *within district office* distribution.

The analysis of variance technique assumes that the impact of a judge on sentencing outcomes is randomly and normally distributed within each district courthouse such that the judge effect has mean = 0 and variance = σ^2 .¹⁷² For instance, suppose that there are 10 judges in a district courthouse. If the judges were identical in their sentencing preferences, and cases are randomly assigned to judges, there would be no impact of a particular judge on sentencing outcomes. Each judge would sentence in the exact same way and variation in the judge effect, as measured by σ^2 , would equal 0. To the extent that judges do differ in their sentencing practices based on personal ideologies or goals,

to a measure of the variance in a judge-specific random variable. Second, the magnitude of an R-squared cannot be taken literally without some discussion of its statistical significance, which is proxied by the linear regression model's significance. Scott's linear regression models are often statistically insignificant, suggesting that judge fixed effects are a poor predictor of sentencing outcomes, but he does not qualify the magnitudes of the R-squared measures. For instance, the model is statistically insignificant in two out of three of the studied periods in Table 2, and four out of five of the studied periods in Table 3, at 34-40. In contrast, measures of inter-judge variance in an analysis of variance can be rigorously tested for statistical significance. Scott acknowledges this issue, noting that "[t]he fact that the model for the *Kimbrough/Gall* period is not significant reinforces the need for caution in interpreting the results for cases not governed by a mandatory minimum.... Although the relationship in the *Kimbrough/Gall* period is strongly positive, the model falls well short of statistical significance." Scott, *supra* note X, at 34-35 FN 177.

¹⁷²See Appendix B for details on the empirical methodology.

one would observe a distribution of judge effects, as measured by the variance or standard deviation in judge effects, σ . The greater the inter-judge variation in outcomes, the larger the σ .

Analysis of variance allows one to estimate the standard deviation of judge effects on sentence length, σ , after controlling for case and defendant characteristics. A finding that $\sigma = 5$ implies a defendant who is assigned to a judge that is one standard deviation “harsher” than the average judge receives a 5 month longer sentence. In order to capture changes in inter-judge disparity, this paper measures σ in periods before *Booker* and after *Booker*. An increase in σ after *Booker* implies an increase in inter-judge sentencing disparity. In particular, I separate the sample timeframe of 2000-2009 into four main periods: (1) *Koon* (October 2000-April 2003), (2) PROTECT Act (May 2003 - January 2005),¹⁷³ (3) *Booker* (January 2005 - November 2007), and (4) *Kimbrough/Gall* (December 2007 - September 2009).¹⁷⁴

3.4. Results on Inter-judge and Regional Disparities

3.4.1. Sentence Length

The following graphs present boxplots of judge average sentence length relative to the average sentence length of the district courthouse in which the judge sits.¹⁷⁵ The box plot captures the spread between the 75 percentile and 25 percentile mean judge departure, as well as outliers. The first panel of Figure 4 indicates that over 50% of judges are sentencing within a few months of the average courthouse mean, with some outliers in both directions. However, the spread of the distributions over the four time periods indicates an increase in the distribution of judge average sentence lengths relative to the court mean following *Kimbrough/Gall*. The spread between the 25th and 75th percentile expands modestly but noticeably across the time periods, as well as the inter-quartile range represented by the box plot whiskers. Following *Kimbrough/Gall*, the number of outliers also increases.

Note, however, that some of the inter-judge disparities may be attributable to uneven applications of mandatory minimums by prosecutors. The second panel of Figure 3.4 presents distributions of average judge sentences for those cases in which a mandatory minimum was not charged, approximately two-thirds of all cases. These cases better represent disparities more likely attributable to judicial behavior. As the figure shows, following *Kimbrough* and *Gall*, judge sentence lengths begin to depart more radically from court averages, with substantially more outlier sentence lengths on both sides of the distribution.

¹⁷³The PROTECT Act became effective as of April 30, 2003.

¹⁷⁴*Booker* was decided on January 12, 2005, and *Kimbrough/Gall* were decided on December 10, 2007. Although the USSC data only contains information on sentencing month and year, rather than date after 2004, the data is coded to denote which January 2005 cases were pre and post *Booker* and which December 2007 cases were pre and post *Kimbrough/Gall*.

¹⁷⁵case composition likely varies across district courts. Thus, I use a measure of judge sentence length that is comparable across all districts, which can be accomplished by calculating average sentence length by judge relative to the mean district courthouse sentence.

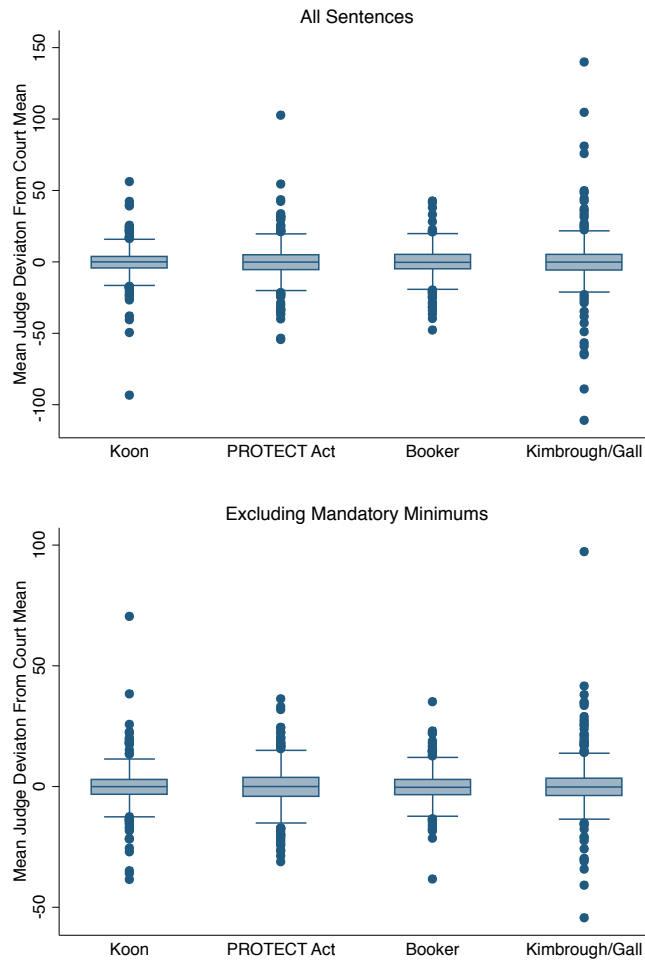


FIGURE 3.4. AVERAGE JUDGE SENTENCE LENGTHS IN MONTHS

Notes: Data is from the random sample 2000-2009.

Statistical analysis of variance confirms the graphical patterns. Table 3.1 presents a measure of σ for sentence lengths, the causal impact of being randomly assigned a 1 standard deviation “harsher” judge in the sentencing district courthouse.¹⁷⁶ Each measure of σ is also accompanied by a 95% confidence interval, which indicates the statistical probability that the true measure of σ lies within the interval range. During the *Koon* period in which the Guidelines were binding and judges were governed by an abuse of discretion standard of appellate review, a defendant assigned to a “harsher” judge received a 2.6 month longer sentence than similar defendants sentenced by an average judge in the courthouse. By the time of the PROTECT Act, a defendant randomly assigned to a harsher judge received almost a 4 month longer sentence. Inter-disparities increased further following *Booker*. *Booker* and *Kimbrough/Gall* induce almost a doubling of inter-judge disparity compared to the *Koon* period. A harsher judge sentenced a defendant 4.6 months longer than the court average in the immediate aftermath of *Booker* and over 5.2 months longer after

¹⁷⁶Table 1 analysis includes all defendants that received a prison sentence, excluding those who received probation.

Kimbrough and Gall.

The second panel of Table 3.1 excludes from the analysis those cases in which a mandatory minimum was charged. During *Koon*, a 1 standard deviation “harsher” judge sentenced a defendant to 1.6 months more than the court average, and 3.2 months longer during the PROTECT Act. Interestingly, inter-judge disparity for non-mandatory minimum cases falls to 2.4 months during *Booker*, rising back up to 3.3 months after *Kimbrough and Gall*. Changes in σ are not significant from the PROTECT Act to *Kimbrough and Gall*, suggesting that on average, inter-judge disparities in sentence lengths of non-mandatory minimum cases have not changed starkly during this period. However, inter-judge disparities are significantly larger following *Gall/Kimbrough* compared to the *Koon* period, more than doubling. This evidence suggests that even in cases in which mandatory minimums are not charged, inter-judge disparities have increased significantly.

Interestingly, the estimates of σ in the bottom panel are almost halved compared to those presented in the top panel where all sentences are included. This finding suggests that a large proportion of inter-judge disparities is driven by the disparate application of mandatory minimums by prosecutors.

TABLE 3.1. INTER-JUDGE VARIATION IN SENTENCE LENGTHS

Period	ALL SENTENCES			
	σ	Lower bound	Upper bound	No. Obs.
Koon	2.616	1.951	3.509	45407
PROTECT Act	3.975	3.118	5.068	38316
Booker	4.651	3.923	5.513	68129
Kimbrough/Gall	5.282	4.519	6.175	48605

Period	EXCLUDING MANDATORY MINIMUMS			
	σ	Lower bound	Upper bound	No. Obs.
Koon	1.599	1.115	2.293	30193
PROTECT Act	3.217	2.633	3.931	27571
Booker	2.392	1.903	3.008	46160
Kimbrough/Gall	3.296	2.783	3.902	33732

Notes: Data is from the random sample from 2000-2009. All regressions contain demographic controls, and controls for offense type, offense level and criminal history. Regressions also contain sentencing year fixed effects, sentencing month fixed effects, and district courthouse fixed effects.

Table 3.2 presents evidence of inter-judge variation in the decision to incarcerate, disentangling the decision to incarcerate from the sentence length decision.¹⁷⁷ Given that *Booker* rendered the Guidelines advisory, a judge could potentially impose no prison sentence on a defendant, even if the Guidelines recommended minimum was non-zero. Indeed, Table 3.2 reveals that inter-judge disparity has increased throughout the four time periods, and most significantly following *Kimbrough and Gall*. During *Koon*, a 1 standard deviation “harsher” judge was 2.9% more likely to incarcerate than the courthouse average. The effect increased to 3.3% during the PROTECT Act, 3.5% during *Booker* and almost 5% following *Kimbrough/Gall*.

¹⁷⁷I define incarceration as a binary indicator, where 1 indicates that the defendant has received a sentence, and 0 indicates no sentence imposed. Defendants who do not received a prison sentence often pay fines and serve probationary periods.

TABLE 3.2. INTER-JUDGE VARIATION IN INCARCERATION RATE

Period	σ	ALL SENTENCES		No. Obs.
		Lower bound	Upper bound	
Koon	.0291	.0239	.0353	51122
PROTECT Act	.0335	.0271	.0414	41713
Booker	.0349	.0297	.0412	73782
Kimbrough/Gall	.0499	.0433	.0575	52586

Notes: Data is from the random sample from 2000-2009. All regressions contain demographic controls, and controls for offense type, offense level and criminal history. Regressions also contain sentencing year fixed effects, sentencing month fixed effects, and district courthouse fixed effects.

3.4.2. Below Range Departures

Figure 3.5 shows the boxplots of average rates of below range departures by judge, relative to the district courthouse mean, for all incarcerated defendants. While the rate of below range departures in the aggregate was lower during the PROTECT Act period (Figure 3.2), the distribution of judge below range departure rates does not appear to be significantly different between *Koon* and the PROTECT Act. In fact, there are far fewer outliers during the first two years following *Booker*. However, inter-judge deviations from the court mean expand visibly following *Kimbrough/Gall*. Figure 3.5 suggests that increasing inter-judge disparities in sentence length as described in Part IV.A are partly attributable to growing inter-judge disparities in the rate of below range departures.

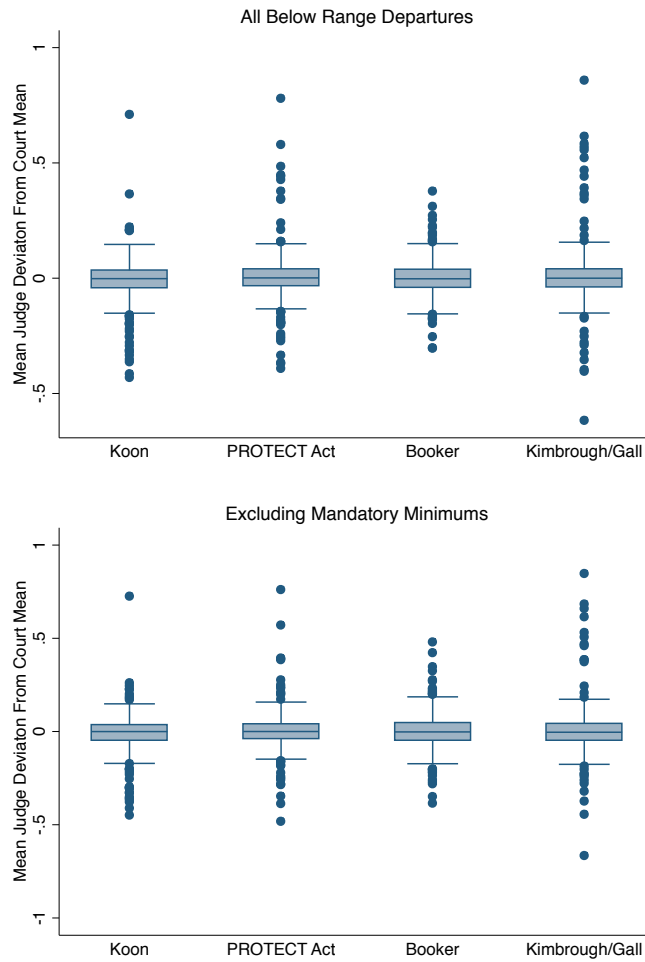


FIGURE 3.5. AVERAGE JUDGE RATES OF BELOW RANGE DEPARTURES

Notes: Data is from the random sample 2000-2009.

Table 3.3 confirms these graphical trends. The top panel of Table 3.3 presents results including all sentences. During the *Koon* period, a defendant who was assigned to a judge 1 standard deviation more “lenient” than the average judge was 4.8% more likely to be sentenced below the Guidelines recommended minimum.¹⁷⁸ During the PROTECT Act, a similar judge was 4.2% more likely to sentence below range. Following *Booker*, the “lenient” judge’s practices deviated more greatly from the courthouse average, with a 5.4% rate immediately following *Booker* and 6.7% rate after *Kimbrough/Gall*. The increased likelihood of below range departures following *Kimbrough/Gall* is statistically significant from the *Koon*-era rate, revealing markedly higher inter-judge disparities.

Excluding cases with mandatory minimums reveals a very similar trend, with the 1 standard deviation more “lenient” judge 5.1% more likely to sentence below range during *Koon*, rising to 7.1% following *Kimbrough/Gall*. Note that the magnitudes of σ when all sentences are included (top panel), and when mandatory minimums are excluded

¹⁷⁸Here, I define a judge who sentences defendants as greater rates below range as more “lenient.” Leniency is used solely to connote lower sentence length.

(bottom panel), are very similar. This finding suggests that inter-judge disparities in below range departures are real and substantial, and not the mere product of mandatory minimums. If anything, measures of inter-judge disparity are lower in most periods when mandatory minimums are *included*. Recall that a mandatory minimum that exceeds the Guidelines recommended minimum trumps the Guidelines minimum, becoming the statutorily binding minimum, thus potentially reducing inter-judge disparity. The findings suggest that the application of mandatory minimums may actually increase the appearance of inter-judge *consistency*.¹⁷⁹

TABLE 3.3. INTER-JUDGE VARIATION IN BELOW RANGE DEPARTURES

Period	σ	ALL SENTENCES		No. Obs.
		Lower bound	Upper bound	
Koon	.0477	.0396	.0576	44338
PROTECT Act	.0427	.0337	.0542	36613
Booker	.0538	.0459	.0632	64781
Kimbrough/Gall	.0668	.0576	.0774	45267

Period	σ	EXCLUDING MANDATORY MINIMUMS		No. Obs.
		Lower bound	Upper bound	
Koon	.0511	.0415	.0629	29369
PROTECT Act	.0522	.0411	.0662	26080
Booker	.0500	.0405	.0618	43390
Kimbrough/Gall	.0712	.0599	.0845	30793

Notes: Data is from the random sample from 2000-2009. All regressions contain demographic controls, and controls for offense type, offense level and criminal history. Regressions also contain sentencing year fixed effects, sentencing month fixed effects, and district courthouse fixed effects.

Not only do mandatory minimums confound the accurate determination of inter-judge disparities, so do below range departures that are government sponsored. Table 4 analyzes inter-judge variation in only those below range departures that are judicially initiated, rather than the result of a government substantial assistance motion. Table 3.4 indicates that the increasing inter-judge disparities in below range departures evidenced in Table 3.3 is consistent in this subset of departures. Inter-judge disparities increased from 4.3% during *Koon* to 5.9% after *Booker* and over 7.4% after *Kimbrough/Gall*, with the lowest inter-judge disparities during the PROTECT Act. Inter-judge disparities similarly increased throughout the period for the subset of cases not subject to a mandatory minimum, from 4.3% during *Koon* to over 7.4% after *Kimbrough/Gall*. These results indicate that in the subset of departures that are most likely attributable to judicial behavior, the PROTECT Act was not only associated with the lowest aggregate rates of downward departures, but also the lowest inter-judge disparities.

¹⁷⁹ See Scott, *supra* note X at 26 (“mandatory minimums may interfere with accurate assessment of inter-judge sentencing disparity by creating the illusion of inter-judge consistency.”).

TABLE 3.4. INTER-JUDGE VARIATION IN BELOW RANGE DEPARTURES
NON-GOVERNMENT SPONSORED DEPARTURES

Period	σ	ALL SENTENCES		No. Obs.
		Lower bound	Upper bound	
Koon	.0426	.0349	.0519	37449
PROTECT Act	.0321	.0247	.0416	33432
Booker	.0588	.0507	.0682	59386
Kimbrough/Gall	.0743	.0645	.0856	42701

Period	σ	EXCLUDING MANDATORY MINIMUMS		No. Obs.
		Lower bound	Upper bound	
Koon	.0434	.0345	.0547	27051
PROTECT Act	.0369	.0274	.0496	25505
Booker	.0553	.0457	.0669	43230
Kimbrough/Gall	.0742	.0628	.0878	31796

Notes: Data is from the random sample from 2000-2009. All regressions contain demographic controls, and controls for offense type, offense level and criminal history. Regressions also contain sentencing year fixed effects, sentencing month fixed effects, and district courthouse fixed effects.

3.4.3. Above Range Departures

Inter-judge disparities have increased not only in the rate of below range departures, but also the rate of above range departures, which comprise approximately 5% of cases. Figure 3.6 presents the distribution of average rates of above range departures by judge, relative to their district courthouse mean, for all incarcerated defendants. The graphs reveal an expansion the distribution of above range departure rates within district courts, particularly between the 25th and 75th percentile of the distribution. Increased inter-judge deviations are also reflected in the rate of above range departures for cases with no mandatory minimums charged. Although there appear to be fewer extreme outliers following *Kimbrough/Gall*, the spread between the 25th and 75th percentile is visibly larger compared to pre-*Booker* spreads.

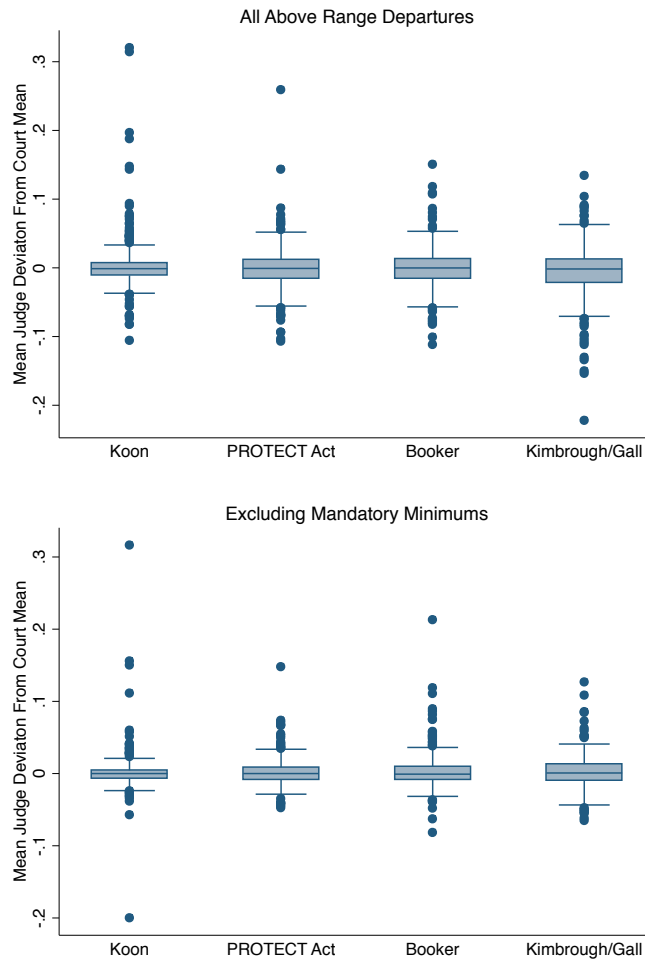


FIGURE 3.6. AVERAGE JUDGE RATES OF ABOVE RANGE DEPARTURES

Notes: Data is from the random sample 2000-2009.

Table 3.5 presents measures of inter-judge variation from the empirical analysis and reveals significant and substantial increases in inter-judge disparities in above range departures. In the top panel where all sentences are analyzed, a 1 standard deviation “harsher” judge was 1.6% more likely to sentence a defendant above range compared to the average judge in the courthouse during *Koon*. While inter-judge variation did not change substantially from *Koon* to the PROTECT Act period, to the first few years after *Booker*, this harsher judge was over 2.7% more likely to sentence a defendant above range after *Kimbrough/Gall*. Inter-judge disparities in above range departures increased by over 70% from the beginning to the end of the time period.

Of course, mandatory minimums explain a sizable fraction of defendants that are sentenced above range when the mandatory minimum trumps the maximum Guidelines recommended sentence. When cases with mandatory minimums are excluded, inter-judge disparities are approximately halved. During *Koon*, inter-judge disparities in above range departures were minimal, with a 1 standard deviation “harsher” judge being only .07% more likely to sentence

above range. However, inter-judge disparities doubled by the end of the time period, to 1.3% after *Kimbrough/Gall*.

TABLE 3.5. INTER-JUDGE VARIATION IN ABOVE RANGE DEPARTURES

Period	σ	ALL SENTENCES		No. Obs.
		Lower bound	Upper bound	
Koon	.0161	.0129	.0202	45184
PROTECT Act	.0182	.0134	.0247	38322
Booker	.0173	.0135	.0222	68120
Kimrough/Gall	.0276	.0227	.0335	48596

Period	σ	EXCLUDING MANDATORY MINIMUMS		No. Obs.
		Lower bound	Upper bound	
Koon	.0067	.0038	.0117	30002
PROTECT Act	.0102	.0072	.0144	27571
Booker	.0139	.0106	.0181	46144
Kimrough/Gall	.0130	.0093	.0183	33734

Notes: Data is from the random sample from 2000-2009. All regressions contain demographic controls, and controls for offense type, offense level and criminal history. Regressions also contain sentencing year fixed effects, sentencing month fixed effects, and district courthouse fixed effects.

3.4.4. Sentencing Practices by Judge Demographics

The previous section finds that inter-judge disparities in sentence length, below range departures, and above range departures have increased significantly following *Booker*, in particular after *Kimrough/Gall*. In this section, I analyze whether increases in inter-judge disparities are idiosyncratic, resulting from all judges changing their behavior in similar ways, or if judges are systematically differing from their colleagues based on observable traits.¹⁸⁰ Recall that due to the random assignment of cases to judges within a district courthouse, any difference in judge sentencing practices can be solely attributable to a judge effect.

To analyze the differential sentencing practices of certain types of judges, I use ordinary least squares regression. The methodology captures how judges differ in their treatment of similar defendants in response to increased judicial discretion, compared to other judges within the same district courthouse. Because cases are randomly assigned to judges within a district court, judge identifiers allow one to compare judges within the same court, capturing judge differences in sentencing rather than different caseloads.

I identify the sources of increasing inter-disparities post *Booker* and post *Kimrough/Gall* using a specification of the form:

$$\begin{aligned}
 Y_{ijkdtm} = & \beta_0 + \alpha_1 * Judge_i * Booker + \beta_0 * Booker + \alpha_2 * Judge_i * Kimrough \\
 & + \beta_0 * Kimrough + \beta_1 * Race_i + \beta_3 * \mathbf{X}_i + Guide_{ijk} + Offtype_i \\
 & + \gamma_d + \delta_t + \gamma_d * \delta_t + \lambda_m + \epsilon_{ijkdtm}
 \end{aligned} \tag{7}$$

¹⁸⁰ See Appendix C.

Y_{ijkdtm} is a sentencing outcome for defendant i , with criminal history category j and offense level k , sentenced in district court d in year t and month m . Main outcomes include sentence length measured in months, and a binary indicator for below range sentencing (sentence length less than the prescribed Guidelines minimum).

$Judge_i$ includes judicial demographics such as race, gender, political affiliation, an indicator for pre vs. post Guidelines appointment, tenure under the Guidelines, and an indicator for pre vs. post *Booker* appointment. The main coefficients α_1 and α_2 capture the impact of particular judicial characteristics on sentencing outcomes in the wake of *Booker*. *Booker* is an indicator variable for defendants sentenced after the *Booker* decision but before *Kimbrough/Gall*. *Kimrough/Gall* is an indicator variable for defendants sentenced after the *Kimrough/Gall*.

$Race_i$ is a dummy variable for defendant i 's race: white, black, Hispanic, or other. \mathbf{X}_i comprises a vector of demographic characteristics of the defendant including gender, age, age squared, educational attainment (less than high school, high school graduate, some college, college graduate), number of dependents, and citizenship status.

$Guide_{ijk}$ includes dummy variables for criminal history category j and offense level k , and each unique combination of criminal history category and offense level. The interaction captures differential sentencing tendencies at each unique cell of the Guidelines grid (258 total). To proxy for underlying offense seriousness and all aggravating and mitigating factors, I control for final offense level. I also control for final criminal history category. $Offtype_i$ is a dummy variable for offense type.

The specification also includes district court fixed effects (γ_d), sentencing year fixed effects (δ_t), and sentencing month fixed effects (λ_m). All standard errors are clustered at the district courthouse level to account for serial correlation.

Consistent with previous research, I find significant and systematic differences in the sentencing practices of both Democratic judicial appointees compared to their Republican appointed peers, and female judges compared to male judges.¹⁸¹ These differences magnified in the aftermath of *Booker* and *Kimrough/Gall*, suggesting that they are some of the sources of the growing inter-judge disparities identified earlier. The coefficients presented in Table 3.6 represent the sentencing tendency of a particular type judge compared to his or her colleagues *within the same district courthouse*, for an identical defendant and case, sentenced in the same month-year.

¹⁸¹ See *supra* notes 28-32.

TABLE 3.6. SENTENCING PRACTICES BY JUDGE CHARACTERISTICS
ALL SENTENCES

	(1) Sentence	(2) Below Range	(3) Below Range Non-Govt	(4) Above Range
Post Booker	1.010 (1.439)	0.0211 (0.0240)	0.0462** (0.0207)	0.0242*** (0.00911)
Tenure	-0.0556 (0.0446)	0.000529 (0.000909)	0.000291 (0.000887)	-0.000543* (0.000276)
Tenure*Booker	0.0461 (0.0614)	-0.000567 (0.00110)	0.000213 (0.000947)	0.00102** (0.000432)
Tenure*Kimbrough	0.0126 (0.0672)	0.00112 (0.00189)	0.00233 (0.00196)	0.000968* (0.000503)
Democratic	-0.522 (0.438)	0.0109 (0.00930)	0.00658 (0.00822)	0.00159 (0.00270)
Democratic*Booker	-0.609 (0.571)	0.00842 (0.00825)	0.00856 (0.00806)	0.00191 (0.00371)
Democratic*Kimbrough	-0.756 (0.731)	0.0223* (0.0115)	0.0269** (0.0120)	-0.00269 (0.00509)
Female	0.537 (0.468)	-0.00247 (0.0126)	-0.00731 (0.0132)	0.00513 (0.00420)
Female*Booker	-1.692*** (0.503)	0.0155 (0.0121)	0.0164 (0.0138)	-0.0100* (0.00538)
Female*Kimbrough	-0.401 (0.576)	0.0110 (0.0192)	0.0145 (0.0203)	0.00118 (0.00553)
Black	-0.873 (0.581)	0.0192 (0.0161)	0.0266 (0.0173)	-0.00818* (0.00438)
Black*Booker	-0.624 (0.951)	-0.0106 (0.0176)	-0.0104 (0.0206)	0.00657 (0.00706)
Black*Kimbrough	0.459 (1.405)	-0.0334 (0.0262)	-0.0367 (0.0315)	0.0137 (0.00946)
Pre Guide	0.561 (0.831)	-0.0266** (0.0134)	-0.0218 (0.0135)	0.00444 (0.00533)
Pre Guide*Booker	-0.765 (0.952)	0.0254 (0.0179)	0.00914 (0.0176)	-0.0137** (0.00690)
Pre Guide*Kimbrough	-0.418 (1.224)	0.0129 (0.0308)	-0.00272 (0.0313)	-0.000320 (0.00995)
Booker	0.276 (0.735)	0.00399 (0.0139)	-0.00635 (0.0132)	-0.00567 (0.00503)
Kimbrough	-0.443 (1.193)	0.00376 (0.0191)	-0.00757 (0.0182)	-0.00170 (0.0124)
Observations	206,292	205,160	178,150	205,160
R-squared	0.785	0.217	0.281	0.083

Notes: Data is from the random sample from 2000-2009. All regressions contain controls for offense type, and dummies for each offense level and criminal history combination. Regressions also contain district courthouse fixed effects, sentencing year and sentencing month fixed effects, and standard errors are clustered at the district courthouse level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Column 1 presents results for sentence length. Female judges significantly altered their practices from their male counterparts within the same courthouse. Immediately after *Booker*, female judges sentenced observably similar defendants to approximately 1.7 months less than their male colleagues.¹⁸² Columns 2 presents results for the rate of below range departures. Inter-disparities in rates of below range departures appear to be somewhat attributable to disparities by judge political affiliation. Following *Kimbrough/Gall*, Democratic judicial appointees are significantly more likely to depart downwards from the Guidelines recommended range, compared to their Republican appointed colleagues. For a similar defendant and crime, Democratic judges were 2.2% more likely to depart downwards. Interestingly, pre-Guidelines (1987) appointees are significantly *less likely* to depart downwards from the Guidelines throughout the entire 2000-2009 period. Inter-judge differences by demographics are also prominent for the subset of below range departures not sponsored by the government, as seen in column 3. Democratic judicial appointees are 2.7% more likely to depart downwards compared to their Republican colleagues. Judges appointed post *Booker* are almost 5% more likely to depart downwards compared to pre-*Booker* appointees.

Finally, columns 4 presents results above range departures. Following *Booker*, female judges are 1% less likely to sentence above range compared to their male colleagues. Inter-judge differences also appear by judge tenure, defined as number of years of experience sentencing under the mandatory guidelines regime for those judges appointed under the mandatory regime. In general, a judge with greater years of experience under the mandatory regime is significantly less likely to sentence above range, but this pattern reverses in the aftermath of *Booker* and *Kimbrough/Gall*, where judges with greater experience are *more* likely to sentence above range. Black judges are in general .08% less likely to sentence above range compared to white judges, and pre-Guidelines appointees are 1.4% less likely to depart upwards after *Booker*. Also striking are the inter-judge differences generated between judges appointed pre-*Booker* and judges appointed post *Booker*. In general, post *Booker* judicial appointees are 2.4% more likely to sentence above range their pre-*Booker* appointed peers.

Table 3.7 presents the results excluding cases with mandatory minimums. Judge differences by political affiliation of appointing present and gender persist. Following *Kimbrough/Gall*, Democratic appointees issue sentences 0.8 months shorter than their Republican colleagues for observably similar defendants, are 3.1% more likely to depart downwards, and 3.1% more likely to depart downwards in cases without government substantial assistance motions. Similarly, female judges issue 0.8 month shorter sentences than their male colleagues following *Booker*. Differences between pre-*Booker* and post *Booker* appointees also remain. Post *Booker* are approximately 5% more likely to depart downwards in all cases, and over 7% more likely when there is no substantial assistance motion.

¹⁸²The *Booker* indicator here represents only the period 2005-2007 prior to *Kimbrough*.

TABLE 3.7. SENTENCING PRACTICES BY JUDGE CHARACTERISTICS
EXCLUDING MANDATORY MINIMUMS

	(1) Sentence	(2) Below Range	(3) Below Range Non-Govt	(4) Above Range
Post Booker	-1.010 (0.683)	0.0496** (0.0242)	0.0713*** (0.0228)	0.00334 (0.00610)
Tenure	-0.0278 (0.0272)	2.90e-05 (0.000791)	-4.93e-05 (0.000825)	-0.000105 (0.000216)
Tenure*Booker	0.0260 (0.0343)	-0.000331 (0.00103)	-0.000152 (0.000976)	0.000196 (0.000287)
Tenure*Kimbrough	-0.0130 (0.0454)	0.00217 (0.00194)	0.00294 (0.00213)	0.000580 (0.000496)
Democratic	-0.382 (0.267)	0.00770 (0.00880)	0.00403 (0.00859)	0.00106 (0.00217)
Democratic*Booker	-0.291 (0.290)	0.00472 (0.0101)	0.00695 (0.0104)	-0.00233 (0.00281)
Democratic*Kimbrough	-0.830* (0.456)	0.0310** (0.0141)	0.0307** (0.0152)	-0.00300 (0.00436)
Female	0.242 (0.324)	-0.00189 (0.0145)	-0.00771 (0.0156)	0.00187 (0.00231)
Female*Booker	-0.811** (0.407)	0.0145 (0.0158)	0.0168 (0.0168)	-0.00334 (0.00325)
Female*Kimbrough	-0.217 (0.501)	0.00493 (0.0219)	0.0123 (0.0221)	0.00155 (0.00472)
Black	-0.501 (0.425)	0.0297* (0.0175)	0.0309* (0.0180)	-0.00315 (0.00215)
Black*Booker	0.220 (0.637)	-0.0258 (0.0200)	-0.0227 (0.0223)	0.00791* (0.00446)
Black*Kimbrough	0.587 (0.811)	-0.0533* (0.0318)	-0.0473 (0.0380)	0.000907 (0.00454)
Pre Guide	0.256 (0.400)	-0.0233 (0.0152)	-0.0229 (0.0157)	0.00367 (0.00330)
Pre Guide*Booker	-0.312 (0.566)	0.0166 (0.0195)	0.0145 (0.0200)	-0.00194 (0.00495)
Pre Guide*Kimbrough	0.364 (0.949)	-0.00734 (0.0331)	-0.0137 (0.0370)	-0.00167 (0.00907)
Booker	0.492 (0.372)	-0.00622 (0.0145)	-0.0174 (0.0138)	0.00414 (0.00351)
Kimbrough	0.184 (0.670)	-0.0201 (0.0189)	-0.0277 (0.0192)	-0.00302 (0.00690)
Observations	141,647	141,386	131,417	141,386
R-squared	0.819	0.244	0.283	0.037

Notes: Data is from the random sample from 2000-2009. All regressions contain controls for offense type, and dummies for each offense level and criminal history combination. Regressions also contain district courthouse fixed effects, sentencing year and sentencing month fixed effects, and standard errors are clustered at the district courthouse level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Overall, these results suggest that sentencing differences associated with judge gender and political affiliation are magnified after *Booker* and/or *Kimbrough/Gall*. Such inter-judge differences are likely sources of growing inter-judge disparities. Given these large changes in inter-judge disparities following *Booker*, judges do not appear to be completely “anchored” to the Guidelines.¹⁸³

However, the finding that post *Booker* judicial appointees are more likely to depart downwards from the Guidelines than pre *Booker* appointees is consistent with a story in which judges with *no* prior experience sentencing under the Guidelines regime are less anchored.¹⁸⁴ The “anchor” of the Guidelines sentence may be more prominent to pre-*Booker* appointees because these judges are more acculturated and experienced under the Guidelines. In contrast, the “anchor” is less prominent for post *Booker* appointees. These potential anchoring differences between pre and post *Booker* appointees suggests that defense lawyer James Felman’s predictions may be true - that disparities may “increase as the years go by and the bench is filled with individuals who have no history with binding guidelines.”¹⁸⁵ Yet, inter-judge disparities that are due to the entrance of new judges to the federal bench might only reflect a short-term surge in disparity. As time goes on and all judges have no history with binding Guidelines, inter-judge disparities attributable to this source may fall.

3.4.5. Regional Disparity: Inter-District Variation

Commentators have suggested that different political climates across districts and circuits can affect sentencing practices,¹⁸⁶ yielding empirical findings that jurisdictional effects are prominent in federal sentencing.¹⁸⁷ The recent 2012 Commission report finds that rates of non-government sponsored below range sentences increasingly depend upon the district court in which the defendant is sentenced and the influence of the Guidelines on sentence length varies significant by circuit court.¹⁸⁸ However, some researchers have found that between-district variation in the

¹⁸³Of course, a degree of anchoring may be occurring, which indicates that these results are only lower bound estimates on increases in inter-judge disparities in a system in which sentencing does not begin with the Guidelines calculation.

¹⁸⁴In robustness checks, I find that the behavior of post *Booker* appointees in my data is not due to the fact that they were George W. Bush appointees based on comparisons with pre-*Booker* George W. Bush appointees. Rather, sentencing behavior seems to be associated with lack of experience under the binding Guidelines.

¹⁸⁵See Felman, *supra* note X, at 98-99.

¹⁸⁶See, e.g., Nora Demlietner, *The Nonuniform Developments of Guideline Law in the Courts*, 6 FED. SENT. REP. 239 (1994) (describing district and circuit specific “personas” in sentencing case law); Jeffrey T. Ulmer & John Kramer, *Court Communities Under Sentencing Guidelines: Dilemmas of Formal Rationality and Sentencing Disparity*, 34 CRIMINOLOGY 383, 402-403 (1996) (Based on an analysis of three county courts in Pennsylvania, the authors argue that local courts operate under formal sentencing standards articulated by a guidelines regime and substantive, extralegal factors relevant to local courts, such as “perceptions of the defendant’s characteristics, local concerns, and court actors’ organizational and individual interests.”); Jeffrey T. Ulmer, *SOCIAL WORLDS OF SENTENCING: COURT COMMUNITIES UNDER SENTENCING GUIDELINES* (1977).

¹⁸⁷See Celesta Albonetti, *Sentencing Under the Federal Sentencing Guidelines: Effects of Defendant Characteristics, Guilty Pleas, and Departures on Sentence Outcomes for Drug Offenses, 1991-1992*, 31 LAW & SOC’Y REV. 789, 815-16 (1997) (finding significant circuit-specific sentencing practices for black and white defendants); Ronald Everett & Roger Wojtkiewicz, *Difference, Disparity, and Race/Ethnic Bias in Federal Sentencing*, 18 J. QUANTITATIVE CRIM. 189 (2002) (finding harsher sentencing in the southern circuits compared to other circuits); Paula Kautt, *Location, Location, Location: Interdistrict and Inter circuit Variation in Sentencing Outcomes for Federal Drug-Trafficking Offenses*, 19 JUSTICE QUARTERLY 633, 659 (2002) (“despite the federal system’s congressionally mandated return to determinate sentencing, extra-legal factors (specifically jurisdictional effects) continue to influence the federal sentencing system and its outcomes directly and indirectly”).

¹⁸⁸United States Sentencing Commission, *REPORT ON THE CONTINUING IMPACT OF UNITED STATES V. BOOKER ON FEDERAL SENTENCING* (2012).

effects of *extralegal* factors on sentencing have not increased following *Booker*.¹⁸⁹

Recall that the identification of the impact of *Booker* on inter-judge disparity within a district courthouse relies on the random assignment of cases to judges. Such random assignment does not exist between districts, such that differences in district sentencing practices are most likely also due to differing caseloads. For instance, the Commission has noted that simple comparisons of regional variations might be attributable to different types of crimes within the general offense categories, such that frauds sentenced in the Southern District of New York are substantially different from frauds sentenced in the District of North Dakota.¹⁹⁰

While I cannot control for unobservable differences across districts, the empirical methodology in this Article does control comprehensively for observable offender and case characteristics.¹⁹¹ For the inter-district results, I utilize the full sample described in Section 3 as random assignment is no longer a prerequisite. In the context of inter-district disparities, analysis of variation now yields an estimate of the standard deviation of *district* effects on sentence length, σ , after controlling for case and defendant characteristics. Thus, a finding of $\sigma = 5$ now suggests that a defendant sentenced in a 1 standard deviation “harsher” district is sentenced to 5 more months in prison, than if he/her were sentenced in an average district court.

Figure 3.7 presents raw distributions of sentence lengths by circuit court, excluding life sentences.¹⁹² While uncontrolled differences cannot be treated as regional effects because districts have very different case compositions, the raw data reveals substantial differences in sentence length, both in the distribution between the 25th percentile and 75th percentile, and presence of outliers.

¹⁸⁹ See Jeffery T. Ulmer et. al, *The “Liberation” of Federal Judges’ Discretion In the Wake of the Booker/Fanfan Decision: Is there Increased Disparity and Divergence Between Courts?*, 28 JUSTICE QUARTERLY 799 (2011).

¹⁹⁰ United States Sentencing Commission, FIFTEEN YEARS OF GUIDELINES SENTENCING: AN ASSESSMENT OF HOW WELL THE FEDERAL CRIMINAL JUSTICE SYSTEM IS ACHIEVING THE GOALS OF SENTENCING REFORM (Nov. 2004) at 99-100 (“Similarly, variations in the rates of a particular type of departure among different districts must be evaluated within a larger context of each district’s distinctive adaptation to the guidelines system. Inferring unwarranted disparity from uncontrolled comparisons of average sentences or rates of departure may be erroneous.”).

¹⁹¹ Nevertheless, the results on inter-district variation should be interpreted with some caution to the extent that there are unobserved differences across district courts that cannot be captured.

¹⁹² Life sentences are top coded as 470 months in the dataset.

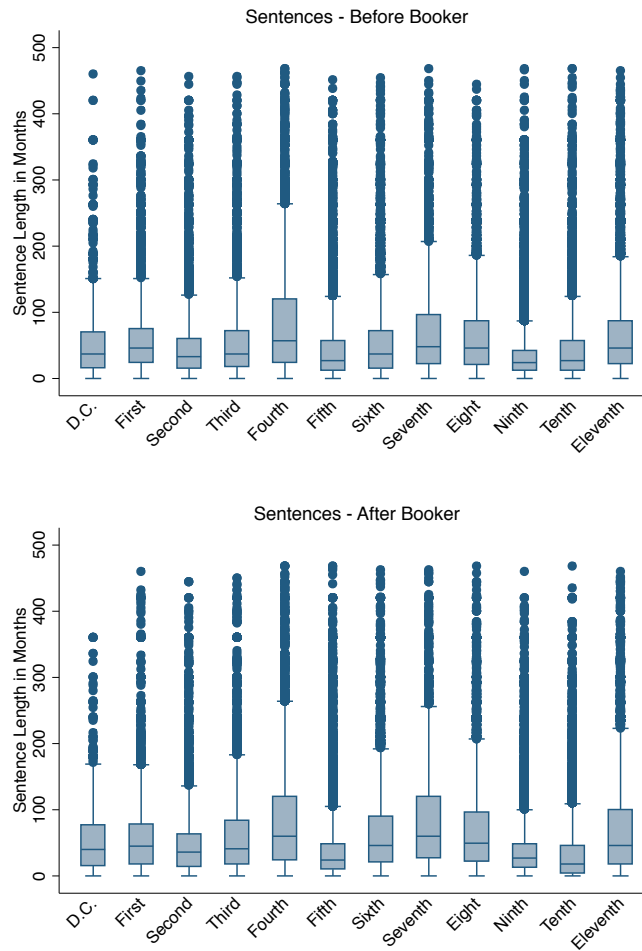


FIGURE 3.7. DISTRIBUTION OF SENTENCE LENGTHS, BY CIRCUIT COURT

Notes: Data is from the full sample 2000-2009.

Table 3.8 shows that after controlling for case and defendant characteristics, there is substantial variation in the sentence that a defendant would receive depending on which district court he is sentenced in. During the *Koon* period, a defendant sentenced in a 1 standard deviation “harsher” district court received a 7.8 month longer prison sentence. This inter-district disparity increased to 8.4 months during the PROTECT Act, to 10.4 months immediately following *Booker*, reaching an 11.3 month difference after *Kimbrough/Gall*. By late 2007, inter-district disparities were significant larger than existed under *Koon*.

Analyzing the subset of cases in which a mandatory minimum was not charged more than halves the magnitude of σ , the measure of inter-district variation. The lower panel of Table 3.8 indicates that a 1 standard deviation “harsher” district court sentenced a defendant to 3.6 months longer than the average district court, 4.4 months longer after the PROTECT Act, 4.9 months longer after *Booker* and 5.2 months longer after *Kimbrough/Gall*. Once again, inter-district variation is statistically greater after *Kimbrough/Gall* compared to *Koon*. Nevertheless, the finding the the magnitude

of inter-district variation is reduced by over half when a statutory minimum is not charged indicates that the application of mandatory minimums is a large contributor to inter-district disparities, particularly in light of the fact that mandatory minimums represent only approximately one-third of the cases.

TABLE 3.8. INTER-DISTRICT VARIATION IN SENTENCE LENGTHS

Period	ALL SENTENCES			
	σ	Lower bound	Upper bound	No. Obs.
Koon	7.799	6.701	9.077	159163
PROTECT Act	8.439	7.232	9.849	83829
Booker	10.397	8.961	12.063	148560
Kimbrough/Gall	11.262	9.692	13.087	106033

Period	EXCLUDING MANDATORY MINIMUMS			
	σ	Lower bound	Upper bound	No. Obs.
Koon	3.578	3.057	4.188	104917
PROTECT Act	4.403	3.758	5.159	58104
Booker	4.920	4.229	5.724	97628
Kimbrough/Gall	5.157	4.419	6.018	72036

Notes: Data is from the full sample from 2000-2009. All regressions contain demographic controls, and controls for offense type, offense level and criminal history. Regressions also contain sentencing year fixed effects, sentencing month fixed effects.

Table 3.9 reveals that district courts also significantly differ in their rates of below range departures. A defendant sentenced in a 1 standard deviation more “lenient” district is 12.1% more likely to be sentenced below the guidelines range, compared to the average district court, during *Koon*. This measure of inter-district variation for below range departures remains relatively constant throughout the entire sample, both including and excluding mandatory minimums. *Booker* and *Kimbrough/Gall* do not appear to have dramatically increased inter-district disparity with regards to downward departures.

TABLE 3.9. INTER-DISTRICT VARIATION IN BELOW RANGE DEPARTURES

Period	σ	ALL SENTENCES		No. Obs.
		Lower bound	Upper bound	
Koon	.1208	.1041	.1402	157103
PROTECT Act	.1198	.1031	.1392	83453
Booker	.1341	.1156	.1555	147774
Kimbrough/Gall	.1289	.1109	.1497	105535

Period	σ	EXCLUDING MANDATORY MINIMUMS		No. Obs.
		Lower bound	Upper bound	
Koon	.1125	.0969	.1308	103720
PROTECT Act	.1113	.0956	.1296	58066
Booker	.1251	.1078	.1453	97568
Kimbrough/Gall	.1256	.1073	.1463	71994

Notes: Data is from the full sample from 2000-2009. All regressions contain demographic controls, and controls for offense type, offense level and criminal history. Regressions also contain sentencing year fixed effects, sentencing month fixed effects.

3.4.6. Prosecutorial Contributions to Disparities

Prosecutors likely contribute to observed inter-judge disparities through their charging behavior. One area of great prosecutorial discretion is the decision to charge an offense that carries a mandatory minimum. Justice Breyer has stated that mandatory minimum statutes “transfer sentencing power to prosecutors, who can determine sentences through the charges they decide to bring.”¹⁹³ Strategic charging of mandatory minimums are likely more prominent after *Booker* as some prosecutors charge mandatory minimums in order to narrow a judge’s discretion.¹⁹⁴

In a 2011 Congressional report on mandatory minimum penalties, the Sentencing Commission found significant variation in the extent which prosecutors applied enhancements for mandatory minimum penalties under drug trafficking offenses.¹⁹⁵ The report documented over 75% of eligible defendants receiving the statutory mandatory minimum penalty in some districts, but none of eligible defendants in other districts receiving the enhancement.¹⁹⁶ Furthermore, recent work by researchers shows evidence of significant racial disparities in prosecutorial charging.¹⁹⁷

Prosecutors are also in charge of the decision to reduce sentences below the mandatory minimum if the defendant offers “substantial assistance” during another investigation or prosecution.¹⁹⁸ If the government files a motion for substantial assistance for a case involving a mandatory minimum sentence, the court has the power to impose a

¹⁹³*Harris v. United States*, 536 U.S. 545, 571 (2002) (Breyer, J., concurring).

¹⁹⁴See Testimony of Patrick J. Fitzgerald, U.S. Attorney, Northern District of Illinois, to the United States Sentencing Commission, at 252 (Sept. 2009) (“[A] prosecutor is far less willing to forego charging a mandatory minimum sentence when prior experience shows that the defendant will ultimately be sentenced to a mere fraction of what the guidelines range is.”).

¹⁹⁵United States Sentencing Commission, REPORT TO CONGRESS: MANDATORY MINIMUM PENALTIES IN THE FEDERAL CRIMINAL JUSTICE SYSTEM, at 252-261 (Oct. 2011).

¹⁹⁶*Id.* at 111-113 (prosecutors reported wide variations in the district practices on seeking statutory minimum penalties).

¹⁹⁷See M. Marit Rehani & Sonja B. Starr, *Racial Disparity in Federal Criminal Charging and Its Sentencing Consequences*, University of Michigan Law & Economics Working Paper 2012. Using data on 58,000 federal criminal cases from 2007-2009, the authors find significant racial disparities in severity of initial charges. In particular, they find that black offenders are on average more than two times as likely to be subjected to a mandatory minimum sentence compared to similar white offenders, and that a major part of the racial gap in sentence length can be attributed to the prosecutorial bias in initial charge.

¹⁹⁸18 U.S.C. §3553(e). A judge has some leeway in reducing sentence length for certain drug trafficking offenses under the “safety valve” provision, which allows a judge to reduce the punishment for low level, first time offenders. See 18 U.S.C. §3553(f). The Report to Congress: Mandatory Minimum Penalties in the Federal Criminal Justice System, *supra* note X, states that in recent years, white defendants in drug cases are more frequently granted the safety valve exception compared to other defendants.

sentence as low as probation.¹⁹⁹ Scholars have commented that the substantial assistance departure provision affords prosecutors immense discretion over both plea bargaining and sentencing outcomes under the Guidelines.²⁰⁰

I find that the application of mandatory minimums appears to be a large contributor to inter-judge disparities. Given the random assignment of cases to judges within a district courthouse, equal application of mandatory minimums among eligible cases would result in no significant judge differences in the rate of mandatory minimums applied. However, empirical results in Table 3.10 reveal small, but significant differences in the percentage of cases with applicable mandatory minimums across judges. A judge 1 standard deviation out in the distribution was 2.3% more likely to see a case with a mandatory minimum during *Koon*, but 3.5% more likely after *Kimbrough/Gall*. The increase in the differential rates of mandatory minimums after *Kimbrough/Gall* coincide with substantial increases in inter-judge disparities in below range departures. These results are consistent with a story in which prosecutors are attempting to rein in judicially induced downward departures through the use of mandatory minimums.

TABLE 3.10. INTER-JUDGE VARIATION IN MANDATORY MINIMUMS

Period	σ	ALL SENTENCES		No. Obs.
		Lower bound	Upper bound	
Koon	.0230	.0180	.0293	51077
PROTECT Act	.0188	.0130	.0272	41697
Booker	.0242	.0196	.0298	73706
Kimbrough/Gall	.0345	.0289	.0411	52551

Notes: Data is from the random sample from 2000-2009. All regressions contain demographic controls, and controls for offense type, offense level and criminal history. Regressions also contain sentencing year fixed effects, sentencing month fixed effects, and district courthouse fixed effects.

TABLE 3.11. INTER-JUDGE VARIATION IN SUBSTANTIAL ASSISTANCE

Period	σ	ALL SENTENCES		No. Obs.
		Lower bound	Upper bound	
Koon	.0372	.0307	.0450	49812
PROTECT Act	.0314	.0248	.0399	41298
Booker	.0286	.0233	.0351	73592
Kimbrough/Gall	.0362	.0301	.0434	52408

Notes: Data is from the random sample from 2000-2009. All regressions contain demographic controls, and controls for offense type, offense level and criminal history. Regressions also contain sentencing year fixed effects, sentencing month fixed effects, and district courthouse fixed effects.

Table 3.12 and 3.13 confirm that a large portion of inter-district differences in the sentencing of observably similar defendants arises from district variation in both the charging of mandatory minimums and the application of a substantial assistance motion. Appendix Table 1 reveals that a defendant sentenced in a 1 standard deviation “harsher” district is approximately 6% more likely to be charged with a mandatory minimum. Appendix Table 2 also presents

¹⁹⁹ According to the Sentencing Commission, substantial assistance motions reduce the average defendant’s sentence length by 50%.

²⁰⁰ See Nagel & Schulhofer, *supra* note X at 550 (“The use of the section 5K1.1 substantial-assistance motion varies from jurisdiction to jurisdiction....There is no limit on the amount of reduction once the motion is submitted. The section 5K1.1 motion is also used to avoid guideline ranges or mandatory minimum sentences for sympathetic defendants – even when there has been no genuine substantial assistance.”); Michael H. Tonry, SENTENCING MATTERS (1996); Jeffrey Standen *Plea Bargaining in the Shadow of the Guidelines*, 81 CALIFORNIA L. REV. 1471 (1993).

evidence of large inter-district differences in the rates of substantial assistance motions, with a defendant being approximately 9% more likely to be granted this form of downward departure in more “lenient” districts. As previously noted, the application of a substantial assistance motion is often applied “to avoid guideline ranges or mandatory minimum sentences for sympathetic defendants - even when there has been no genuine substantial assistance.”²⁰¹ However, inter-district differences in average rates of mandatory minimums and rates of substantial assistance motions do not appear to have increased significantly following *Booker* and *Kimbrough/Gall*.

TABLE 3.12. INTER-DISTRICT VARIATION

APPLICATION OF MANDATORY MINIMUMS				
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0571	.0491	.0663	183732
PROTECT Act	.0652	.0559	.0760	94434
Booker	.0584	.0502	.0680	165139
Kimbrough/Gall	.0662	.0568	.0771	117536

Notes: Data is from the full sample sample from 2000-2009. All regressions contain demographic controls, and controls for offense type, offense level and criminal history. Regressions also contain sentencing year fixed effects, and sentencing month fixed effects.

TABLE 3.13. INTER-DISTRICT VARIATION

APPLICATION OF SUBSTANTIAL ASSISTANCE				
Period	σ	Lower bound	Upper bound	No. Obs.
Koon	.0868	.0747	.1009	176736
PROTECT Act	.0909	.0782	.1058	92736
Booker	.0904	.0779	.1049	163978
Kimbrough/Gall	.0789	.0679	.0918	117328

Notes: Data is from the full sample sample from 2000-2009. All regressions contain demographic controls, and controls for offense type, offense level and criminal history. Regressions also contain sentencing year fixed effects, and sentencing month fixed effects.

3.5. Policy Recommendations

This section describes three of the major proposals for reform of federal sentencing after *Booker*. I describe each in turn, and then apply the empirical findings in this paper to shed light on the various proposals.

3.5.1. “Topless” Guidelines System

Within a few months after *Booker*, the Department of Justice recommended a new “topless” guidelines system, in which judges would be bound by the Guidelines minimum, but not the maximum.²⁰² Echoing the topless guidelines regime first proposed by Professor Frank Bowman, this construction would still allow judicial fact-finding to facts that raised the minimum applicable sentence, remaining constitutional under the principles espoused in *Blakely*.²⁰³ Recall

²⁰¹Nagel & Schulhofer, *supra* note X, at 550.

²⁰²See Gonzales, *supra* note X, at 326 (favoring “the construction of a minimum guideline system”).

²⁰³See Memorandum from Frank Bowman to U.S. Sentencing Comm’n (June 27, 2004), 16 FED. SENTENCING REP. 364, 367 (2004).

that *Blakely* applied the Sixth Amendment to challenge judicial fact-finding which raised a defendant's maximum sentence.²⁰⁴ As a result, a "topless" guidelines system would comport with both *Blakely* and the Court's holding in *Harris v. United States* that facts triggering a mandatory minimum sentence could be found by a judge.²⁰⁵ However, while a topless regime addresses potential Sixth Amendment raised in *Blakely* and *Booker*, it does not limit judicial discretion to sentence above the applicable guidelines maximum. Moreover, the topless regime takes the prior mandatory guidelines as the baseline, which some argue "would constitute a step backwards in the development and evolution of the federal sentencing system by exacerbating some of the worst features of the pre-*Booker* federal sentencing."²⁰⁶ By binding judges to the applicable minimum sentence, the topless "fix" would likely re-introduce concerns of prosecutorial power in their charging and plea bargaining decisions.²⁰⁷

The empirical evidence from Part IV reveals that a "topless" guidelines proposal, while comporting with the Sixth Amendment, would plausibly aggravate disparities that are attributable to prosecutor charging decisions. Table 1 and Table 5, which present evidence of inter-judge disparities and inter-district disparities, are reduced by almost a factor of two when mandatory minimums are excluded from analyses. These results suggest that the decision to charge a mandatory minimum contributes substantially to inter-judge differences in sentence length, such that these decisions are not made equally across all eligible cases. The results also indicate that mandatory minimum practices may differ largely across U.S. district courts. As a result, any proposal that binds judges to the applicable minimum sentence would ascribe greater power to prosecutors, likely resulting in inequitable disparities. Moreover, results in Table 6 suggest that there have been substantial increases in inter-judge disparities in above range departures even when a mandatory minimum is not charged. As a result, to the extent that a "topless" regime seeks to limit judicial discretion, it does so in an asymmetric manner.

3.5.2. "Blakely-ized" Guidelines

Justice Breyer's ultimate remedy for the Sixth Amendment issues facing the federal sentencing guidelines was to declare the guidelines "effectively advisory." But one could have imagined another approach to "*Blakely*-ize" the Guidelines. Indeed, the Justices dissenting from the Breyer remedial opinion in *Booker* suggested leaving the mandatory guidelines intact, but requiring that aggravating facts triggering longer maximums be proven by a jury

²⁰⁴*Blakely*, 542 U.S. 296.

²⁰⁵536 U.S. 545, 567-69 (2002). However, some commentators have suggested that the Court's holding in *Harris* may not survive after *Booker*. See, e.g., Erwin Chemerinsky, *Making Sense of Apprendi and its Progeny*, 37 MCGEORGE L. REV. 531, 541 (2006); Frank O. Bowman III, *Mr. Madison Meets a Time Machine: The Political Science of Federal Sentencing Reform*, 58 STAN. L. REV. 235, 261 (2005). The Court has recently granted review in a case presenting this exact question of whether the Court's decision in *Harris* should be overruled. *Alleyne v. United States*, 133 S. Ct. 420 (2012).

²⁰⁶See Berman, *Tweaking Booker*, *supra* note X, at 363. Berman also discusses potential constitutional challenges to a "topless" guidelines system.

²⁰⁷*Id.* at 364 ("Consequently, the most problematic facets and the most disconcerting consequences in terms of prosecutorial power, disparity, and evasion experienced in the pre-*Booker* federal sentencing system would likely be aggravated by the enactment of any sort of topless guideline *Booker* fix.").

beyond reasonable doubt, or admitted by the defendant.²⁰⁸

However, introducing jury fact-finding into a mandatory guidelines system is likely particularly complex. Justice Breyer in his *Booker* remedial opinion mused over how jury fact-finding might work, asking “[w]ould the indictment have to allege, in addition the elements of robbery, whether the defendant possessed a firearm, whether he brandished or discharged it, whether he threatened death...?”²⁰⁹ Additionally, to the extent that the mandatory guidelines regime contributed to prosecutorial discretion and disparity, jury fact-finding in the face of extensive plea bargaining “would move the system backwards in respect to both tried and plea-bargained cases” by effectively “prohibit[ing] the judge from basing a sentence upon any conduct other than the conduct the prosecutor chose to charge.”²¹⁰ Other scholars have echoed the concern that *Blakely*-izing the current version of the Guidelines would be procedurally unworkable and overwhelm juries required to make findings of fact.²¹¹ Addressing some of Justice Breyer’s concerns, a 2005 American Bar Association Report suggested a version of the *Blakely*-ized system espoused by Justice Stevens in his *Booker* dissent, accompanied with “simplifying the guidelines by reducing both the number of offense levels and the number of adjustments and presenting the remaining, more essential, culpability factors to the jury.”²¹²

This Article cannot comment on the relative abilities of judges and juries to determine the applicability of aggravating and mitigating factors. Even supposing that juries are capable of fact determinations of aggravating and mitigating factors, under the “*Blakely*-ized” guidelines, once a jury has made factual determinations as to conduct based on what a prosecutor chose to charge, a judge is bound by these determinations. For instance, if a jury did not make a factual determination with respect to a potential mitigating factor, such as acceptance of responsibility by the defendant, a judge would not be allowed to consider this factor, even if it were applicable. This Article provides evidence suggesting that a large component of disparities stem from prosecutorial charging decisions. To the extent that the disparities introduced by the prosecutorial charging behavior are increased by jury fact-finding in a mandatory guidelines system, disparities may become exacerbated.

3.5.3. Sessions Proposal

Former Sentencing Commission Chair Judge Sessions has recommended a simplification of the Guidelines to provide for fewer and broader sentencing ranges within a presumptive regime, with enhancing facts to be charged in

²⁰⁸*Booker*, 543 U.S. at 284-84 (Stevens, J., dissenting) (“Rather than engage in a wholesale rewriting of the SRA, I would simply allow the Government to continue doing what it has done since this Court handed down *Blakely*—prove any fact that is *required* to increase a defendant’s sentence under the Guidelines to a jury beyond a reasonable doubt.”).

²⁰⁹*Booker*, 543 U.S. at 254.

²¹⁰*Booker*, 543 U.S. at 256 (“plea bargaining would likely lead to sentences that gave greater weight, not to real conduct, but rather to the skill of counsel, the policies of the prosecutor, the caseload, and other factors that vary from place to place, defendant to defendant, and crime to crime. ...plea bargaining of this kind would necessary move federal sentencing in the direction of diminished, not increased, uniformity in sentencing”). For a more thorough discussion of the potential problems with this particular recommendation, see Berman, *Tweaking Booker*, *supra* note X, at 365-71.

²¹¹See Bowman, *Beyond Band-Aids*, *supra* note X, at 191 (“the consensus view is that the Guidelines as now written are simply too complex and confusing to operate through juries”).

²¹²ABA Criminal Justice Section, Report and Recommendation on *Booker* (Jan. 2005), reprinted in 17 FED. SENT. REP. 335, 339 (2005).

an indictment and proved to a jury beyond a reasonable doubt.²¹³

Sessions argues in favor of a new sentencing regime that balances two goals: (1) the need to reduce unwarranted sentencing disparities curbing the ability of judges to use subjective notions of justice to mete out punishment, and (2) giving judges discretion to tailor sentencing to the unique circumstances of offenders and offenses.²¹⁴ Thus, Sessions proposes a return to a more simplified presumptive guidelines that affords judges discretion within broader sentencing ranging, subject to fewer mandatory minimum statutes.²¹⁵ Sessions recommends a reduction in the number of possible sentencing ranges, but broader sentencing ranges associated with offense levels and criminal history category combinations, to afford judges greater discretion.²¹⁶ In order to comply with the constitutional requirements identified in *Blakely*, Sessions suggests that any facts that would increase the base offense level to increase the applicable maximum sentence would have to be proven by a jury beyond a reasonable doubt, unless admitted to by the defendant.²¹⁷

Sessions also proposes simplifying the Guidelines by reducing the number of aggravating or mitigating factors that increase or decrease the base offense level under Chapter Two and Chapter Three of the Guidelines Manual, which many have argued are overly complex.²¹⁸ In deciding which aggravating factors to keep within the sentencing guidelines, Sessions argues in favor of the strategy suggested by Justice Breyer - empirically reviewing which enhancements in Chapter Two are commonly used.²¹⁹

Finally, Sessions suggests a new form of appellate scrutiny because “[t]he threat of reversal [on appeal] is a key component of [effective] guidelines,” with within range sentences “essentially unreviewable on appeal ... [unless] a district court refused to consider all relevant factors or instead considered a prohibited factor, such as a defendant’s race or gender.”²²⁰ In contrast, Sessions proposes “relatively strict scrutiny by the appellate court” for downward departures.²²¹

Critiques of the Sessions proposal argue that the proposal would eliminate “judicial feedback to the Commission and constructive evolution of the guidelines would virtually cease” as judges would have limited authority in setting the applicable sentence range, and sentencing outside the range.²²² As a result, both scholars and district court judges have expressed that the current advisory guidelines best achieves the goals of sentencing. Commentators have sug-

²¹³William K. Sessions III, *At the Crossroads of the Three Branches: The U.S. Sentencing Commission’s Attempts to Achieve Sentencing Reform in the Midst of Inter-Branch Power Struggles*, 26 J.L. & POL. 305, 346-50 (2011).

²¹⁴Sessions, *supra* note X, at 339.

²¹⁵*Id.* at 340.

²¹⁶*Id.* at 340-45 (describing recommended changes to the current Guidelines sentencing chart).

²¹⁷*Id.* at 346.

²¹⁸*Id.* at 347-48; see also Frank O. Bowman III, *The Failure of the Federal Sentencing Guidelines*, 105 COLUM. L. REV. 1315, 1341 (2005); Stephen Breyer, *Federal Sentencing Guidelines Revisited*, 11 FED. SENTENCING REP. 180 (1999) (“[T]he Guidelines are simply too long and too complicated.”).

²¹⁹Sessions, *supra* note X, at 349; Stephen Breyer, *supra* note X, at 11 (“... I believe the Commission should review the present Guidelines, acting forcefully to diminish significantly the number of offense characteristics attached to individual crimes. The characteristics that remain should be justified for the most part by data that shows their use by practicing judges to change sentences ...”).

²²⁰Sessions, *supra* note X, at 353-54.

²²¹*Id.* at 353-54 (“District courts’ choices of sentences within the applicable cells on the grid would be essentially unreviewable on appeal so long as the courts considered all of the relevant aggravating and mitigating factors identified in the application notes and all other relevant factors in the Guidelines Manual before imposing a particular sentence.”).

²²²Baron-Evans & Stith, *supra* note X, at 1716.

gested that the current advisory guidelines system should be retained because it reflects the right balance between various actors in federal sentencing,²²³ and approximately 70% of surveyed district judges believe that the guidelines reduced unwarranted sentencing disparity among similarly situated defendants, and increased certainty and fairness in sentencing.²²⁴ Of these district judges surveyed in 2010, over 75% prefer the current advisory guidelines system to other alternatives.²²⁵ In contrast, 14% of judges favored a version of the *Blakely*-ized proposal - “[a] system of mandatory guidelines that comply with the Sixth Amendment and have broader sentencing ranges than currently exist, coupled with fewer statutory mandatory minimums.”²²⁶ Only 3% of judges preferred a return to the pre-*Booker* guidelines system, suggesting that the overwhelming majority of judges are opposed to a return to presumptive guidelines, as proposed by Judge Sessions.²²⁷ Undoubtedly, *Booker* has given judges the freedom to consider the particular circumstances of the offense and traits of the defendant. To the extent that growing inter-judge disparities are reflected of these considerations, disparities are warranted and judicial discretion is desirable. On the other hand, some have suggested that the advisory guidelines have been accompanied by increases in unwarranted disparities.²²⁸

The empirical findings in this Article suggest that inter-judge disparities have increased from the period of mandatory or “presumptive” regime to post *Booker* sentencing. Indeed, a return to “presumptive” guidelines would mechanically reduce inter-judge disparities by greatly limiting judicial discretion. However, the empirical evidence from Part IV seeks to ascertain the effect of sentencing regime on inter-judge disparities in outcomes that are most likely attributable to judge behavior. Differences in sentence lengths can be attributable to both judge disparities as well as differences in charging of mandatory minimums. The findings in this Article suggest that a return to a presumptive regime, without any changes in mandatory minimums, would only go partway in reducing disparities, and curtail potentially desirable judicial discretion.

While this Article does not provide evidence supporting a return to “presumptive” guidelines, it does suggest that strictness of appellate review is a potentially important constraint on judicial discretion in sentencing. Inter-judge disparities in below range departures were generally lowest during the PROTECT Act. Furthermore, empirical evidence suggests that *Booker* alone did not contribute to recent increases in inter-judge disparities. In the first two years after *Booker*, inter-judge disparities are not statistically different from that during *Koon*. Rather, it is the impact of *Booker* plus reduced appellate scrutiny following *Rita*, *Gall* and *Kimbrough* that are responsible for any increases in inter-judge disparities.

²²³Baron-Evans & Stith, *supra* note X, at 1681. *See also* Michael Tonry, *The U.S. Sentencing Commission’s Best Response to Booker is to Do Nothing*, 24 FED. SENT. REP. 387 (2012); Sara Sun Beale, *Is Now the Time for Major Sentencing Reform?*, 24 FED. SENT. REP. 382 (2012).

²²⁴*See* U.S. Sentencing Commission, RESULTS OF SURVEY OF UNITED STATES DISTRICT JUDGES JANUARY 2010 THROUGH MARCH 2010 (June 2010), at 23 (Question 17, Table 17).

²²⁵*See* U.S. Sentencing Commission, RESULTS OF SURVEY OF UNITED STATES DISTRICT JUDGES JANUARY 2010 THROUGH MARCH 2010 (June 2010), at 23 (Question 19, Table 19).

²²⁶*Id.*

²²⁷*Id.*

²²⁸Sessions, *At the Crossroads*, *supra* note X; Bowman, *Nothing is Not Enough*, 24 FED. SENT. REP. 356 (June 2012) (“[T]he post-*Booker* advisory system retains most of the flaws of the system it replaced, while adding new ones, and its sole relative advantage - that of conferring additional (and effectively unreviewable) discretion on sentencing judges - is insufficient to justify its retention as a permanent system.”).

Thus, reforms to strengthen the degree of appellate review could possibly reduce inter-judge sentencing disparities. In *Gall*, the Court did not require appellate courts to insist upon “extraordinary” circumstances to justify a sentence outside the Guidelines recommended range, specifically rejecting stronger justifications for sentences that departed more greatly from the Guidelines.²²⁹ In order to constrain inter-judge disparities, the Commission could require district court judges to provide a heightened justification for more severe departures from the prescribed sentence, without coming too close to an “impermissible presumption of unreasonableness for sentences outside the Guidelines range.....[which] would not be consistent with Booker.”²³⁰

3.6. Conclusion

Exploiting the random assignment of cases to judges in district courthouses representing 73 U.S. district courts, this Article finds a significant increase in inter-judge disparities from the *Koon* period to after *Kimbrough/Gall*. Inter-judge disparities increased in a variety of outcomes: the decision to incarcerate, sentence lengths conditional on incarceration, below range departures, and above range departures. Increased inter-judge disparities persisted even excluding cases in which mandatory minimums were charged, suggesting that judges are not fully anchored to the Guidelines.

Increases in between-judge differences following *Booker* and *Kimbrough/Gall* appear to be linked to observable judicial demographics such as gender, political affiliation of appointing president, and whether a judge has ever sentenced under the mandatory regime. I also find modest evidence of increases in inter-district differences following *Kimbrough/Gall*, with large inter-district differences in sentence length, below range departures, and rates of mandatory minimums. However, the magnitudes of the both inter-judge and inter-district disparities are drastically smaller when mandatory minimums are excluded, suggesting that prosecutorial charging decisions are a major contributor to sentencing disparities.

Overall, these results suggest that the shift to an advisory guidelines regime under *Booker*, coupled with lowered standards of appellate scrutiny, have led to somewhat greater inter-judge variation. Prosecutorial charging decisions, at least in the case of mandatory minimums, appear to play a substantial role in explaining disparities. While a first step in disentangling the sources of disparities ascribable to various actors, a primary limitation of this Article is its inability to thoroughly analyze all the disparities that can arise in earlier stages of the criminal justice systems, such as through charging and plea bargaining. Nevertheless, the results of this Article caution against proposals to move back towards a sentencing system in which judges are bound by the decisions of prosecutors. Instead, the Article suggests that it may be wise to modify standards of appellate review, as well as revisit the desirability of mandatory minimums.

²²⁹*Gall*, 552 U.S. at 47 (“In reviewing the reasonableness of a sentence outside the Guidelines range, appellate courts may therefore take the degree of variance into account and consider the extent of a deviation from the Guidelines. We reject, however, an appellate rule that requires “extraordinary” circumstances to justify a sentence outside the Guidelines range. We also reject the use of a rigid mathematical formula that uses the percentage of a departure as the standard for determining the strength of the justifications required for a specific sentence.”).

²³⁰*Id.* at 47.

4. THE IMPACT OF FEDERAL AND STATE OSHA PROGRAMS ON WORKPLACE SAFETY, WAGES AND EMPLOYMENT

4.1. Introduction

In response to increasing concern about workplace safety and the decentralized and often inefficient state worker's compensation programs, the federal government passed the Occupational Safety and Health Act (OSHA) in 1970. Existing state statutory remedies and common law actions were viewed as inadequate to protect workers from unsafe working conditions. The OSHA Act mandated that private sector employers create safe working environments to prevent work-related injuries, illnesses, and deaths, in compliance with health and safety standards set by the Secretary of Labor. Since 1971, when OSHA was implemented, occupational deaths have been reduced by 62% and workplace injuries lowered by 42%. However, workplace injuries are still a major public policy concern. OSHA estimates that each year there are over 6,000 workplace fatalities and 50,000 deaths from workplace related illnesses, in addition to over 5 million nonfatal work related injuries.

While the statute was promulgated at the federal level, Section 18 of the act permits states to create and enforce their own OSHA programs, as long as the state programs are "at least as effective" as the federal one. States must set job safety and health standards that are comparable to federal standards. In fact, the majority of states adopt standards identical to the federal ones. However, since the state programs must be at least as stringent as the federal OSHA, many states have implemented innovations above and beyond federal OSHA regulation. State regulation of OSHA programs may also be more efficient than federal regulation since state regulators are more aware of the business environment within certain industries in a state.

Additionally, any state plan must cover public sector (state and local government) employees, whereas states that remain under federal mandate cover only private sector employees. The benefit of providing coverage to public sector employees is large, since many hazardous occupations, such as firefighting, corrections, law enforcement, and forms of transportation are in the public sector. State approved OSHA plans cover over 57 million workers, over 40% of the private sector and over 10 million public sector employees (Grassroots, Worker Protection, 2008 OSHSPA Report).

A state must undergo an extensive procedure in order to receive federal OSHA approval for its own plan. First, OSHA must approve its developmental plan, which includes the regulation, enforcement and other logistics of the state program, which must be implemented within three years. Upon completion of the developmental plan, states are eligible for certification. To formally suspend federal enforcement of activities covered by the state plan, the state and OSHA may choose to enter into an "operational status agreement." Another route taken by most states after certification is the final approval, which also ends federal authority over the state's workplace health and safety matters. Although there is no federal enforcement in approved states, the federal OSHA can continue to monitor the

state plans and funds up to 50% of the state's operational costs.

Twenty-two states and jurisdictions currently operate their own state plans, which cover both private and public sector employees. Four other states/territories cover only public employees, Connecticut, New Jersey, New York and the US Virgin Islands. The remaining states are covered by the Federal OSHA, which applies only to private sector employees. Table 4.1 lists the state OSHA programs and dates of initial approval, certification, and final approval/operational status agreement, as well as whether the state program has different standards than the federal program.

The effectiveness of OSHA programs remains a controversial issue to this day. Many state regulated OSHA programs have implemented voluntary programs that provide incentives for workplaces to maintain excellent safety records. However, critics argue that these voluntary programs are inefficient because they target workplaces with strong safety records, providing no incentives for less safe employers to improve the workplace.

A recent New York Times article on OSHA programs claims that OSHA's ultimate enforcement tool is the prosecution of cases. There have been many efforts to make worker deaths caused by the workplace a felony, but politicians from both parties have rejected these efforts. The maximum criminal fine was increased once in 1984 and all civil fines were increased in 1991. However, an analysis done by the Times suggests that these changes have had only modest effects on workplace safety.

Labor groups, journalists, and even OSHA officials believe that federal regulation of OSHA has fallen behind a vast number of state regulated OSHA programs, which have either required notification to a prosecutor upon workplace safety related deaths, or increased criminal penalties. As of today, California is the only state that actively prosecutes employers who kill workers by violating safety laws. Conviction can result up to three years in prison, along with a \$1.5 million penalty. The Times found that California had more safety violation prosecutions than all other states combined, potentially contributing to its low workplace fatality rate, compared to the rest of the country.

Table 4.1. State Statutes

State	Initial Approval	Certified	Final Approval	Operational Status	Different Standards
Alaska	7/31/73	9/9/77	9/28/84		
Arizona	10/29/74	9/18/81	6/20/85		
California	4/24/73	8/12/77		Yes	Yes
Connecticut	10/2/73	8/19/86			
Hawaii	12/28/73	4/26/78	4/30/84		
Indiana	2/25/74	9/24/81	9/26/86		
Iowa	7/20/73	9/14/76	7/2/85		
Kentucky	7/23/73	2/8/80	6/13/85		
Maryland	6/28/73	2/15/80	7/18/85		
Michigan	9/24/73	1/16/81		Yes	Yes
Minnesota	5/29/73	9/28/76	7/30/85		
Nevada	12/4/73	8/13/81	4/18/00		
New Jersey	1/11/01				
New Mexico	12/4/75	12/4/84		Yes	
New York	6/1/84	8/18/06			
North Carolina	1/26/73	9/29/76	12/10/96		
Oregon	12/22/72	9/15/82	5/12/05	Yes	Yes
Puerto Rico	8/15/77	9/7/82		Yes	
South Carolina	11/30/72	7/28/76	12/15/87		
Tennessee	6/28/73	5/3/78	7/22/85		
Utah	1/4/73	11/11/76	7/16/85		
Vermont	10/1/73	3/4/77		Yes	
Virgin Islands	8/31/73	9/22/81	4/17/84		
Virginia	9/23/76	8/15/84	11/30/88		
Washington	1/19/73	1/26/82		Yes	Yes
Wyoming	4/25/74	12/18/80	6/27/85		

Note: Data from OSHA.

Using a differences in differences methodology, I explore the impact of state OSHA regulation on traditional enforcement tools, nonfatal and fatal injury rates, as well as wages and employment.

I find that state regulation of OSHA programs leads to an increased use of inspections per capita and citation of violations per capita. Despite the increased use in these enforcement tools, however, in the more recent period from 1996-, state regulated OSHA programs have no significantly different rate of nonfatal injuries, compared to federally regulated programs. This might reflect laxer enforcement by states today or possibly inefficiencies at the state level.

When I compare state programs with standards more stringent than the federal program, evidence suggests that state regulation is associated with lower fatalities. This suggests that state regulation may be more efficient than federal regulation, and/or that voluntary programs and other state OSHA innovations are effective at promoting workplace safety. A case study on the California civil penalty increase and increase in criminal sanctions reveals that this reform led to a significant fall in the number of fatalities, indicating that greater magnitude of sanctions and prosecution are potentially effective enforcement tools in promoting greater workplace safety.

Finally, I analyze the impact of state regulation of OSHA on wages and employment. The increased use of inspections and issuance of violations following state regulations may have also increased costs for employers of hiring workers. I find that there is a compensating differential for workplace safety, as wages fall significantly following certification of a state regulated OSHA program, particularly in the more dangerous industries. While demand for workers may have fallen, the supply of workers also seems to have shifted out, as workers valued the benefit of greater workplace safety, leading to no change in employment, or an increase in the more dangerous industries. This finding is suggestive that workers valued workplace safety at cost to employers, leading to full shifting of the cost onto wages, with no efficiency loss in hours worked.

4.2. Literature Review

Early empirical studies of the federal OSHA found that enforcement between 1972 and 1975 did not lead to measurable improvements in workplace safety, due to high costs of compliance and ineffective enforcement (Viscusi 1979). Bartel and Thomas (1985) also tracked OSHA compliance and find a low correlation between workplace compliance and injury rates in the same time period. These scholars attribute the failure of OSHA regulation in improving workplace safety to the low penalty associated with violation and the low chance of inspection. Smith (1976) argues that in the initial years of OSHA implementation, there were very reduced incentives for compliance, and safety standards promulgated by the Secretary of Labor were often not highly related to the major causes of workplace injuries.

However, OSHA adopted a new records inspection system in 1981 that sought to better target and keep track of high hazard workplaces, allowing regulators to inspect high risk violators more effectively. In a study that takes into account these changes in record keeping, Ruser and Smith (1988) find that firms subject to records check inspections

had 5-14% reductions in workplace injuries, suggesting that the policy change increased the effectiveness of OSHA. Viscusi (1986) confirms these findings in a study of the incidence of lost workdays due to illness or injury in the manufacturing sector after the implementation of the records check inspection system.

Analyses of OSHA in the 1980s and 1990s also confirmed that OSHA enforcement, in the form of inspections and penalties, was associated with a lower prevalence of workplace injuries, but many have found that these effects, while strong initially, declined over time (Gray and Jones 1991a,b, Gray and Scholz 1994, Gray and Mendeloff 2002). Weil finds that the decline in injury in the construction industry from 1987-1993 was not due to the direct impact of OSHA enforcement (2001). The finding that federal OSHA programs have had, at most, short term impacts on reducing workplace injuries, is surprising given the evidence that despite low expected penalties, workplaces have high rates of compliance (Weil 1996).

More recently, researchers have begun to exploit the different state and federal OSHA programs, although research has treated these programs as homogeneous in all aspects besides enforcement. Using the variation in state and federal enforcement of the construction industry from in the 1980s and 1990s, Morantz (2009) finds that state inspectors use traditional enforcement tools less frequently. She also finds that state enforcement is associated with a lower rate of occupational fatalities, but interestingly, a higher rate of nonfatal injuries, which she suggests may be due to underreporting of nonfatal injuries in federally regulated states, or the differential responses of nonfatal and fatal injuries to different enforcement tools. Similarly, Bradbury (2006) uses data from 1981-1995 and finds that state regulated OSHA programs have lower workplace fatalities than federally regulated programs.

Additionally, perhaps due to the minimal effects of OSHA on injury rates, there has been no study of the impacts of the OSHA health and safety mandates on wages or employment levels, which I explore. Given the low expected penalty associated with violation of OSHA, employers may not comply with the mandate, thus having no effect on wages and employment.

Ultimately, the existing empirical work has shown minimal effects of OSHA on injury rates, and very little work on exploiting variation across state programs. Thus, there has been no consensus on whether OSHA programs have actually increased efficiency and employee welfare.

4.3. Theory

Given the critiques of OSHA as being highly intrusive, bureaucratic, yet altogether ineffective, it is of great interest to explore the actual impacts of OSHA on workplace safety. Many scholars point to minimal staff and resources within OSHA as key to its ineffectiveness, but given limited resources, state innovations such as increases in penalties and voluntary cooperative programs with employers may be more effective at reducing workplace injury than traditional federal enforcement measures. Some legal scholars even argue that OSHA should be abolished, claiming that the financial costs from worker's compensation are enough to provide economic incentives to increase workplace safety

(Kniesner and Leeth 1995, Maukestad and Helm 1989). Ultimately, I explore which types of governance strategies, whether at the federal or state OSHA level, are most effective in reducing workplace injury and illness.

4.3.1. Regulation

Workplace safety regulation falls within a broader regulation literature on legal rules that govern the employer-employee relationship. Other legal rules include workers' compensation programs, employee fringe benefits mandates, wrongful discharge laws, unemployment insurance, and minimum wage laws. The law and economics literature on employment law considers both the theory and the empirics of employment regulations, addressing the efficiency and welfare consequences of these legal interventions. Interventions through employment law are particularly relevant in the face of market failure, such as informational asymmetries and externalities.

In the area of workplace safety, in the absence of market failure, one would expect that less safe working conditions would be offset by higher wages through compensating differentials. But there are often informational failures such that employees are not aware of the risks associated with a certain workplace, and workplace injuries often have externality effects that are not redressed through the market, leading to an oversupply of labor at a given wage. Workplace safety mandates can then in theory improve both efficiency and employee welfare, but is ultimately an empirical question.

4.3.2. Enforcement

Theories behind OSHA compliance and penalty provisions are well situated within a larger literature on public enforcement of law, reviewed by Shavell and Polinsky (2000). At the most basic level, enforcement depends a choice of the probability of sanction, as well as the magnitude of sanctions which affects the behavior of regulated parties. The administrative agency in charge of OSHA has control over a variety of enforcement tools including the number of inspectors, the probability of inspection, the magnitude of violation penalties, and type of penalty (civil or criminal). A priori, there is no reason to expect that states with state regulated OSHA programs would be more effective at enforcement than states with federally regulated OSHA programs.

4.3.3. Federal versus State Enforcement

Previous research has mainly studied the effect of federal OSHA enforcement on inspections and injury rates, while noting that changes in record keeping and other enforcement mechanisms lowered the rate of occupational injuries and illnesses. As mentioned before, very few studies have explored the impact of state level OSHA programs. While Morantz looks at the difference between state and federal enforcement, her analysis begins in the late 1980s, after the majority of states with their own OSHA programs had adopted them. As a result, she is only able to use cross state variation and not variation within state, over time. However, using data from the early 1970s on inspections and

violations which covers the period during which state OSHA programs were initially approved and certified, I plan to study the differential impact of state and federal OSHA programs on the use of these enforcement tools. I am also able to exploit within state variation over time in my analysis of the impact of state OSHA implementation on wages and employment. Data on nonfatal and fatal injuries are not available until after the implementation of state OSHA programs, so these analyses rely only on cross state variation. Figures 4.1, 4.2, 4.3 plot the distribution of initial approval, certification and final approval for state level OSHA programs.

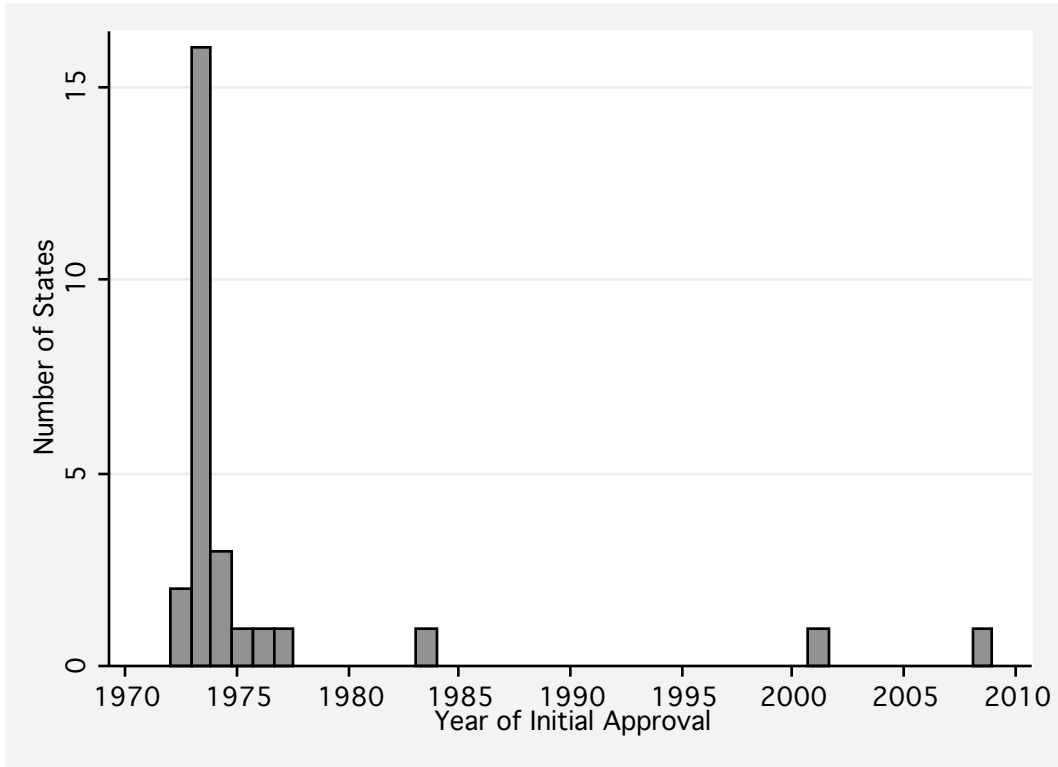


Figure 4.1: Distribution of States by Initial Approval Year

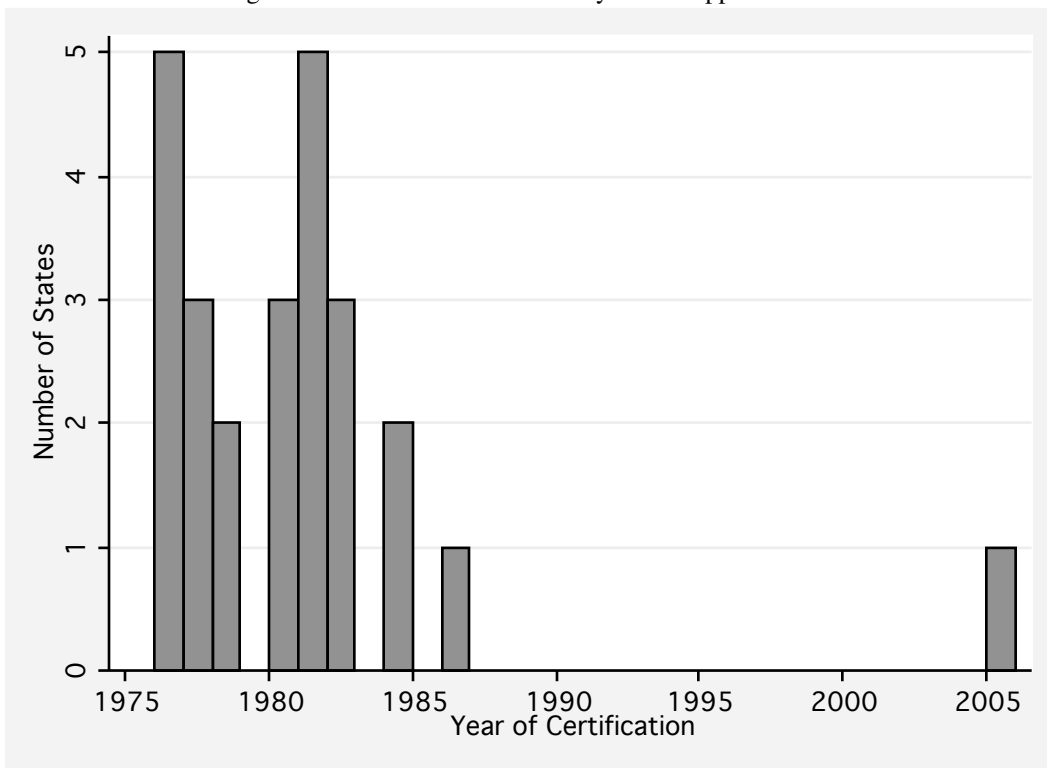


Figure 4.2: Distribution of States by Certification Year

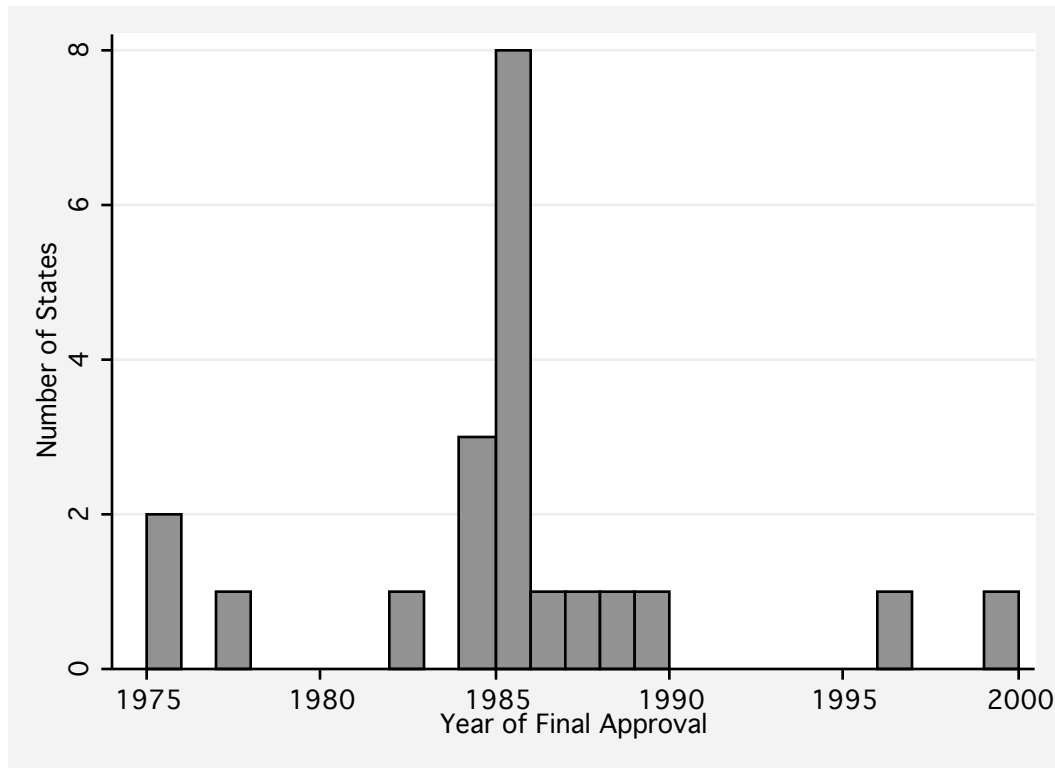


Figure 4.3: Distribution of States by Final Approval Year

Key in the research is the differential impact of federal versus state regulation as studied by Morantz and Bradbury. Bradbury notes that interjurisdictional competition between states, created when states administer and regulate their own OSHA programs, can lead to the optimal level of regulation. Furthermore, state regulators may be more aware of local industry and business conditions. Given more information regarding monitoring costs, state level regulation can certainly be more efficient than federal regulation. Scholz and Gray (1997) argue that OSHA programs should not only increase safety through the use of penalty systems, but also by facilitating greater cooperation between employers and workers. It is likely that this cooperation can be greater when OSHA programs are regulated at the state level. On the other hand, it is important to note that state regulation can also lead to corruption between industry and state regulators from political pressure, leading to “regulatory capture.” (Bradbury 2006). This type of regulatory capture may be stronger at the state than federal level as state OSHA officials may have a greater connection with state industry interests than federal OSHA officials.

However, these studies have treated the state and federal OSHA programs as homogeneous, focusing only on the differential enforcement of these programs. Yet, different state programs were plausibly created for very different purposes. Depending on the political party in power at the time, certain states may have implemented a state OSHA program in order to avoid federal enforcement and regulation. Additionally, many state OSHA programs adopted the federal OSHA mandate almost verbatim and there are no innovations beyond those required in the federal OSHA.

Therefore, to the extent that federal OSHA has been ineffective, it is likely that these states have also had minimal effects on workplace injury rates.

On the other hand, there are also significant differences between some state programs and the federal OSHA. In particular, four states (California, Michigan, Washington and Oregon) have OSHA programs that are much more rigorous than the federal OSHA. It seems plausible that these state level innovations can lead to differences in outcomes of interest, as measured by inspection and violation rates, injury rates, and potentially wages and employment levels of employees.

Thus, a comparison between federal and all state regulated OSHA programs can be potentially misleading. I hypothesize that depending on the political affiliation of a state at the time OSHA was implemented, as well as the degree to which state programs differ from the federal program, there can be two distinct effects of state OSHA regulation. Thus, I study the differential effects of federal and state regulation by disaggregating states by their political climate and their OSHA standards.

4.3.4. Differences in Penalty Structures

Employers who do not comply can be subject to abatement orders or violations, which result in civil penalties. In the original OSHA Act, inspectors who conducted health and safety inspections at workplaces could propose penalties ranging from nothing for nonserious violations, often due to noncompliance with technical requirements, to \$1,000 for serious violations, and up to \$10,000 for willful or repeat violations, issued when employers are not in compliance with safety standards or when a high number of injuries have occurred in the previous period.

One of the most important differences between the federal and state OSHA programs is the penalty structures. Since the inception of the federal OSHA in 1970, there has been only one change in the penalty structure, passed in the 1990 Omnibus Budget Reconciliation Act. The Act significantly increased the penalty maximums for serious, willful and repeat violations, although many critics argue that the penalties and inspection rates are still far too low. The penalties increased seven fold, to penalty maximums of \$7,000 for serious, other-than-serious, failure to abate, and posting violations, \$70,000 for willful and repeat violations, and a \$5,000 floor for willful violations.

Following the federal act, states with approved OSHA programs were required to adopt these federal changes, effective in 1991 and 1992. However, since then, several states have further increased penalties, sometimes even subjecting willful violations to criminal charges.

California is recognized as having the most strictly enforced OSHA program in the nation. In 2000, California passed a new OSHA act that subjected employers to greater civil penalties and more importantly, introduced large criminal penalties for willful violations of health and safety regulations. The definition of serious violation was expanded and civil penalties for a serious violation were increased from \$7,000 to \$25,000. Employers were also newly liable for criminal penalties of \$1,500,000 and imprisonment for three years. The law also was the first time

that public agencies were subject to monetary violations so government entities were no longer exempt from penalties. More importantly, the burden of proof on serious violations was effectively reversed to fall on employers.

4.3.5. Incentive Programs

Additionally, many state OSHA program officials recognize that enforcement and financial penalties are not the only way to incentivize employees to keep workplaces safe. Incentive programs have frequently formed from state level OSHA experimentation. For instance, the most famous incentive program was the Maine 200. Created in 1993, the Maine 200 pilot program asked the 200 most hazardous workplaces in Maine to cooperate with OSHA in developing and implementing a comprehensive health and safety program in exchange for a low rate of future inspections. OSHA officials found that participating companies significantly reduced their injury and illness rates. Following success in Maine and nine other states with similar pilot programs, President Clinton expanded the Maine 200 program nationally in 1995 in his "New OSHA" plan in the form of the Cooperative Compliance Program (CCP). However, industry groups challenged the CCP program and in 1999, the Federal Court of Appeals of the District of Columbia held that the CCP was not properly adopted by OSHA and invalidated the program. Subsequently, OSHA cancelled the program.

Nevertheless, many states have created other forms of voluntary compliance programs for hazardous industries since 1999. Created as an experimental program in California in 1979, and adopted federally by OSHA in 1982, Voluntary Protection Programs (VPP) recognize outstanding participating employers. If a worksite passes the standards of the VPP program in the form of very low injury rates and various other health and safety standards, employers are exempt from routine inspections for three years. States also have Safety and Health Achievement Recognition Programs (SHARP) that offer eligible employers exemptions from programmed inspections. However, the VPP and SHARP programs involve the cooperation between state OSHA program and employers with a history of excellent safety, rather than targeting the most hazardous workplaces.

4.4. Data

I use a variety of data sets for this paper. All data is state panel data, either at the industry or individual level, where industry is characterized as the major division into which an occupational industry falls into. There are 9 industry level divisions: Agriculture, Mining, Construction, Manufacturing, Transportation, Wholesale Trade, Retail Trade, Finance and Services.

I obtained data on inspections and violations through OSHA's Integrated Management Information System (IMIS). The IMIS contains complete records of all inspections (type and scope), violations (type and penalty amount) by state and industry level since 1972. For states that eventually pass their own OSHA programs, IMIS contains federal inspection data up to the year the state was certified.

Data on occupational injury and fatality rates is from the Bureau of Labor Statistics. I obtained data on workplace fatalities from the Census of Fatal Occupational Injuries (CFOI) from 1992 to present. The CFOI contains records of all workplace fatalities broken down by state, industry, and type of fatality, from 2003 to present. Data on nonfatal workplace injuries and illnesses is from the Survey of Occupational Injuries and Illnesses (SOII), which contains state level data from 1976 to present, although I only currently have access to SOII data from 1996-. The SOII contains information on all types of injuries and time away from work by state by industry.

Data on wages and employment levels come from the Current Population Survey (CPS), which begins reporting hourly wages in 1978. Control variables at the industry level are obtained from the CPS. State level demographic variables are obtained from the Statistical Abstract. Data on state workers' compensation programs is from the National Academy of Social Insurance.

When regressions exploit the within-state variation, over time, in the creation of state OSHA programs, where the dependent variable is inspections, violations, or violation rate, I control for number of inspections instigated by employee complaints, unionization rate, industry size and average age of workers at the industry division level. I also control for state population, real per capita income. These are controls that are deemed to be important by previous researchers, such as Weil (1998) and Morantz (2009). See the following Table 4.2 and Table 4.3 for summary statistics of state regulated and federally regulated OSHA programs.

Since data on nonfatal and fatal injuries are only available since the late 1990s, I am only able to exploit cross state variation in state vs. federal OSHA programs since the majority of state regulated OSHA programs had already been implemented at this point. In these regressions, I control for all the industry division characteristics, as well as one year lagged number of inspections and one year lagged number of violations, and state worker's compensation characteristics such as total number of workers covered under workers' compensation, total workers' compensation benefits, and workers' compensation replacement rate, calculated as benefits as a proportion of covered wages.

Table 4.2. Summary Statistics for State OSHA Programs

Variable	1980 Mean	1990 Mean	2000 Mean	SD	Minimum	Maximum
Inspections	58.63	412.57	291.79	580.94	1	10805
Violations	151.59	1082.39	663.13	1367.08	0	23276
Violation Rate	.505	.584	.565	.217	0	1
Employee Complaints	18.02	66.35	59.47	109.71	0	2117
Population (in 1000s)	4794.36	5383.17	6115.74	6468.44	402	35484.45
Real Per Capita Income	128.04	146.32	168.43	26.70	93.55	242.35
Industry % Union		.019	.009	.022	0	1
Industry Size	600.88	444.69	383.61	642.24	0	5907
Industry Average Age	38.72	42.17	42.96	6.57	18	62.24

Note: Data from IMIS, CPS and Statistical Abstract.

Table 4.3. Summary Statistics for Federal OSHA Programs

Variable	1980 Mean	1990 Mean	2000 Mean	SD	Minimum	Maximum
Inspections	241.62	149.34	118.88	440.38	1	9025
Violations	467.74	512.51	231.98	1290.11	0	46756
Violation Rate	.475	.589	.545	.251	0	1
Employee Complaints	59.60	34.97	27.36	81.50	0	1434
Population (in 1000s)	5133.52	4753.83	5334.39	4765.99	519	22118.51
Real Per Capita Income	115.49	138.89	163.54	27.28	85.04	274.23
Industry % Union		.016	.007	.016	0	.5
Industry Size	530.38	441.55	384.99	541.03	0	4148
Industry Average Age	38.82	42.61	43.41	6.74	18	66

Note: Data from IMIS, CPS and Statistical Abstract.

4.5. Empirical Methodology

4.5.1. Impact of State OSHA programs on Inspections and Violations

To analyze the effect of state OSHA programs on inspections and violations, I use a differences-in-differences methodology exploiting cross time and state variation, using states that do not have state OSHA programs as controls. Regressions are variations of the following empirical specification:

$$\text{IndependentVariable}_{i,s,t} = \beta \mathbf{X}_{i,s,t} + \alpha \text{STATELAW}_{s,t} + \sum_i \text{Industry fixed effects}_i + \sum_s \text{State fixed effects}_s + \sum_t \text{Time fixed effects}_t \quad (8)$$

The independent variables of interest are the total number of inspections per capita, violations per capita and the violation rate, defined as the probability of violation, given an inspection. I also run regressions breaking down the total number of inspections into type of inspection, which includes Accident, Followup, Planned, State, Federal, Complete and Partial. In the above specification, i indexes industry, s indexes state and t indexes time during the period 1972-2009. STATELAW is the explanatory variable of interest, and $X_{i,s,t}$ is a vector of both industry level and state level covariates, as described in the previous section. When industry level covariates are used, the data covers the period 1979-2009. Standard errors are clustered at the state level.

The explanatory variable, STATELAW, equals 1 in each year after the state OSHA program was created. I define STATELAW in three ways, based on initial approval, certification and final approval. My definition of choice is certification. Bradbury (2006) suggests that there is no real distinction between certification and federal approval, indicating that after certification, federal enforcement is essentially suspended.

I also disaggregate state OSHA programs to better explore the differences between state OSHA programs. I analyze the impact of state OSHA programs that have different standards from all other states, including those that have identical standards to the federal program. In these regressions, I define STATELAW to equal 1 in each year after a state OSHA program with different standards was created.

I also compare across state OSHA programs by political affiliation of the state, defined as the political party of the governor in power at the time of the creation of the state OSHA program. When I disaggregate states with their own OSHA programs by political affiliation of the state, I define STATELAW equal to 1 in each year after a state OSHA program was created by a Democratic governor, and a separate STATELAW variable for state OSHA programs created by a Republican governor.

4.5.2. Impact of State OSHA programs on Nonfatal and Fatal Injuries

I analyze the effect of state OSHA programs, compared to federal programs, on nonfatal injuries and illnesses, from 1996-2008. Because of a change in industry classification in 2002, from the standard SIC to NAICS system, specific occupations are not directly comparable before and after the change, but at the industry level, the comparisons are more valid. I also study the effect of state OSHA programs on fatalities at the state and industry level, from 2003-2008.

Since nonfatal and fatal injury data is only available post certification of state OSHA programs, I can only compare across states, using states that do not have state OSHA programs as controls. Regressions are variations of the

following empirical specification:

$$\text{Independent Variable}_{i,s,t} = \beta \mathbf{X}_{i,s,t} + \alpha \text{STATELAW}_{s,t} + \sum_i \text{Industry fixed effects}_i + \sum_t \text{Time fixed effects}_t \quad (9)$$

Here, the explanatory variable of interest, STATELAW, equals 1 if a state has its own OSHA program. Similar to the previous subsection, comparison are also conducted by different standards and political climate at the time of implementation.

4.5.3. Impact of Changes in Penalty Structure

To study the impact of the California penalty change in 2000, I use a similar differences in differences methodology to analyze the change in inspections per capita, violations per capita and violation rate following the change. STATELAW is equal to 1 in each year after California adopted the penalty change, using various control groups to test the robustness of the results.

4.5.4. Impact of State OSHA Programs on Wages and Employment

As before, I use a differences-in-differences methodology exploiting cross time and state variation, using states that do not have state OSHA programs as controls. I use individual level observations when exploring the effect of state regulated OSHA programs on wages and employment (as measured by hours worked). Individual covariates that are controlled for are age, sex, race, education, marital status and union coverage.

4.6. Results - Impact of State OSHA programs on Inspections Per Capita and Violations Per Capita

Table 4.4 presents the regression results for the impact of all three stages of implementation of a state OSHA program: initial approval, certification and final approval. It is clear from the results that initial approval of a state OSHA program has no significant effect on total number of inspections, whether one controls for industry level characteristics, state level demographics, or both. However, by the time a state OSHA program is certified, there is a significant positive effect on inspections per capita. Similarly, there is a significant increase in inspections per capita by the time of final approval, although the point estimate is smaller.

Similarly, it appears that certification and final approval of a state OSHA program are both associated with an increase in the total number of violations per capita (Table 4.5). While both inspections and violations increase at the time of certification, the overall violation rate did not change significantly after either initial approval, certification or final approval. See Table 4.6. I will henceforth use certification as the indicator for when state enforcement begins.

Table 4.4. Impact of State OSHA Laws on Inspections Per Capita

	(1) Inspections	(2) Inspections	(3) Inspections	(4) Inspections	(5) Inspections	(6) Inspections	(7) Inspections	(8) Inspections	(9) Inspections
State Law/Initial Approval	-0.00657 (0.0058)			-0.00555 (0.0044)			-0.00626 (0.0058)		
State Law/Certified		0.0717*** (0.026)			0.0712*** (0.026)			0.0723*** (0.026)	
State Law/Final Approval			0.0250** (0.010)			0.0263** (0.011)			0.0259** (0.011)
Employee Complaints	0.000145* (0.000084)	0.000138* (0.000080)	0.000138 (0.000085)	0.000146* (0.000085)	0.000139* (0.000079)	0.000134 (0.000089)	0.000139 (0.000091)	0.000131 (0.000086)	0.000130 (0.000093)
Population	0.00000320 (0.0000026)	0.00000366 (0.0000025)	0.00000206 (0.0000021)				0.00000351 (0.0000031)	0.00000406 (0.0000031)	0.00000244 (0.0000026)
Real Per Capita Income	0.0000424 (0.00034)	-0.0000625 (0.00036)	0.000137 (0.00034)				0.0000400 (0.00035)	-0.0000676 (0.00038)	0.000134 (0.00035)
Industry % Union				-0.116 (0.083)	-0.120 (0.086)	-0.121 (0.082)	-0.115 (0.083)	-0.119 (0.086)	-0.120 (0.082)
Industry Size				0.00000262 (0.0000064)	0.00000340 (0.0000061)	0.00000443 (0.0000069)	0.00000363 (0.0000072)	0.00000467 (0.0000068)	0.00000503 (0.0000075)
Industry Average Age				0.000159 (0.00100)	0.000156 (0.00100)	0.0000598 (0.00096)	0.000211 (0.0011)	0.000182 (0.0011)	0.000129 (0.0010)
State Controls	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Industry Controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9257	9257	9257	11552	11552	11552	9257	9257	9257
R ²	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50

Note: The dependent variable in all regressions is the total number of inspections. Data are industry-state-year observations. State, industry and year fixed effects are included in all specifications.

Standard errors are clustered at the state level. Regressions with state controls have data from 1972-2004, regressions with industry level controls have data from 1979-2009. Coefficients that are significant at the .1, .05, .01 percent level are indicated with *, **, ***, respectively.

Table 4.5. Impact of State OSHA Laws on Number of Violations Per Capita

	(1) Violations	(2) Violations	(3) Violations	(4) Violations	(5) Violations	(6) Violations	(7) Violations	(8) Violations	(9) Violations
State/Initial Approval	0.0184 (0.015)			0.0148 (0.010)			0.0172 (0.016)		
State/Certified		0.144** (0.063)			0.141** (0.065)			0.144** (0.063)	
State/Final Approval			0.0505* (0.026)			0.0525* (0.027)			0.0506* (0.027)
Employee Complaints	0.000398 (0.00025)	0.000383 (0.00024)	0.000383 (0.00025)	0.000423 (0.00025)	0.000409* (0.00024)	0.000399 (0.00026)	0.000408 (0.00026)	0.000391 (0.00025)	0.000390 (0.00027)
Population	0.00000788 (0.0000061)	0.00000886 (0.0000058)	0.00000564 (0.0000052)				0.00000718 (0.0000066)	0.00000830 (0.0000062)	0.00000512 (0.0000057)
RPC Income	-0.000186 (0.00080)	-0.000340 (0.00088)	0.0000613 (0.00079)				-0.000196 (0.00082)	-0.000357 (0.00089)	0.0000418 (0.00081)
Industry % Union				-0.257 (0.18)	-0.266 (0.19)	-0.268 (0.17)	-0.256 (0.18)	-0.265 (0.19)	-0.266 (0.17)
Industry Size				-0.00000928 (0.000016)	-0.00000782 (0.000015)	-0.00000575 (0.000016)	-0.00000697 (0.000017)	-0.00000503 (0.000016)	-0.00000438 (0.000017)
Industry Average Age				-0.00100 (0.0020)	-0.00104 (0.0020)	-0.00123 (0.0019)	-0.000976 (0.0021)	-0.00105 (0.0021)	-0.00115 (0.0020)
State Controls	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Industry Controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9257	9257	9257	9257	9257	9257	9257	9257	9257
R ²	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49

Note: The dependent variable in all regressions is the total number of violations. Data are industry-state-year observations. State, industry and year fixed effects are included in all specifications. Standard errors are clustered at the state level. Regressions with state controls have data from 1972-2004, regressions with industry level controls have data from 1979-2009. Coefficients that are significant at the

.1, .05, .01 percent level are indicated with *, **, ***, respectively.

Table 4.6. Impact of State OSHA Laws on Violation Rate

	(1) Viol Rate	(2) Viol Rate	(3) Viol Rate	(4) Viol Rate	(5) Viol Rate	(6) Viol Rate	(7) Viol Rate	(8) Viol Rate	(9) Viol Rate
State/Initial App	0.0149 (0.027)			0.0187 (0.020)			0.0145 (0.027)		
State/Certified		0.0214 (0.016)			0.0154 (0.014)			0.0213 (0.016)	
State/Final App			0.00250 (0.010)			0.00228 (0.0092)			0.00219 (0.010)
Employee Complaints	0.0000174 (0.000024)	0.0000150 (0.000024)	0.0000164 (0.000024)	0.0000188 (0.000024)	0.0000173 (0.000024)	0.0000178 (0.000024)	0.0000227 (0.000026)	0.0000200 (0.000026)	0.0000217 (0.000026)
Population	0.00000632** (0.0000032)	0.00000648** (0.0000032)	0.00000623* (0.0000033)				0.00000599* (0.0000033)	0.00000618* (0.0000033)	0.00000592* (0.0000033)
RPC Income	0.00000311 (0.00031)	0.00000206 (0.00030)	0.0000405 (0.00031)				0.00000942 (0.00031)	0.00000699 (0.00031)	0.0000443 (0.00031)
Industry % Union				-0.103 (0.095)	-0.105 (0.095)	-0.104 (0.095)	-0.123 (0.11)	-0.124 (0.11)	-0.123 (0.11)
Industry Size				-0.00000205 (0.0000050)	-0.00000210 (0.0000050)	-0.00000219 (0.0000050)	-0.00000363 (0.0000061)	-0.00000340 (0.0000061)	-0.00000358 (0.0000061)
Industry Average Age				-0.000661 (0.00082)	-0.000673 (0.00082)	-0.000685 (0.00082)	-0.0000222 (0.0011)	-0.0000413 (0.0011)	-0.0000388 (0.0011)
Observations	9257	9257	9257	11552	11552	11552	9257	9257	9257
R^2	0.23	0.23	0.23	0.22	0.22	0.22	0.23	0.23	0.23

Note: The dependent variable in all regressions is the violation rate, the percentage of inspected firms in each industry that are charged with at least violation. Data are industry-state-year observations.

State, industry and year fixed effects are included in all specifications. Regressions with state controls have data from 1972-2004, regressions with industry level controls have data from 1979-2009.

Coefficients that are significant at the .1, .05, .01 percent level are indicated with *, **, ***, respectively.

The control covariates also enter in with the expected signs. As expected, an increase in the number of employee OSHA complaints in an industry is somewhat significantly and positively associated with the total number of inspections per capita. An increase in state population is associated with a higher violation rate, or probability of violation given inspection.

Additionally, since I have more years of data when using only industry controls, I use only these in all later regressions. Tables 4.4-4.6 show that the inclusion of state level demographics does not change the regression coefficient by much. All later results are robust to the inclusion of state level controls.

One can also see the significant increase in the mean number of inspections and mean number of violations after certification in Figures 4.4, 4.5, 4.6 and 4.7.

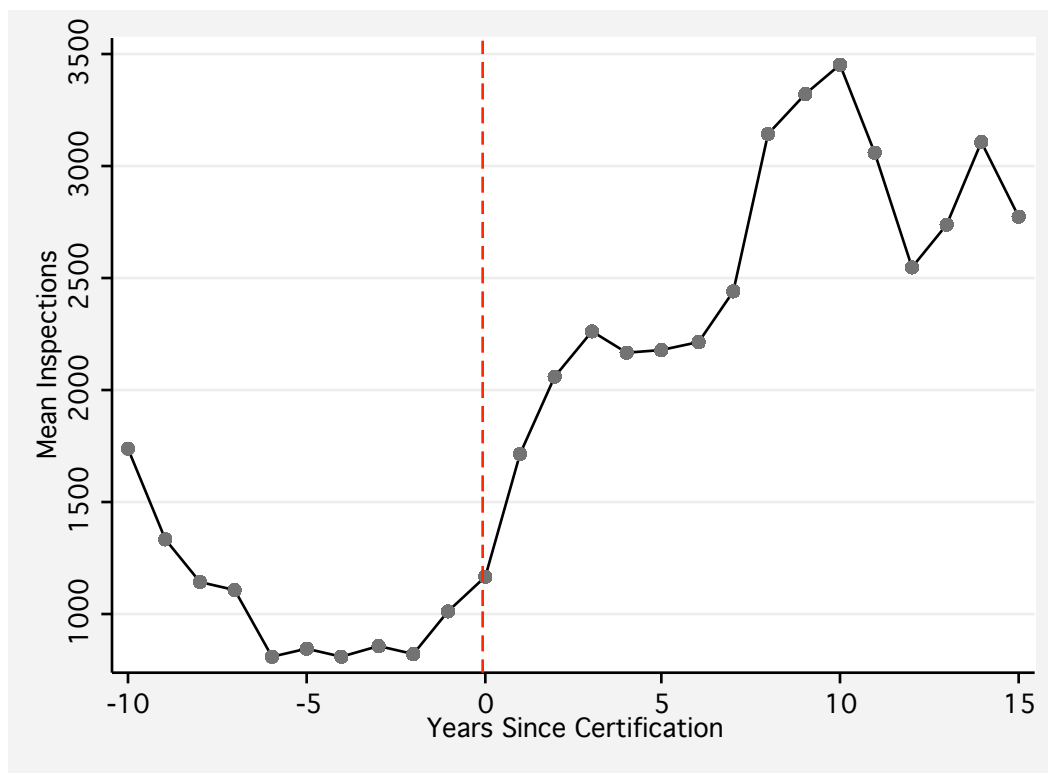


Figure 4.4. Trends in Mean Inspections for State Regulated OSHA Programs

Note: Data from IMIS.

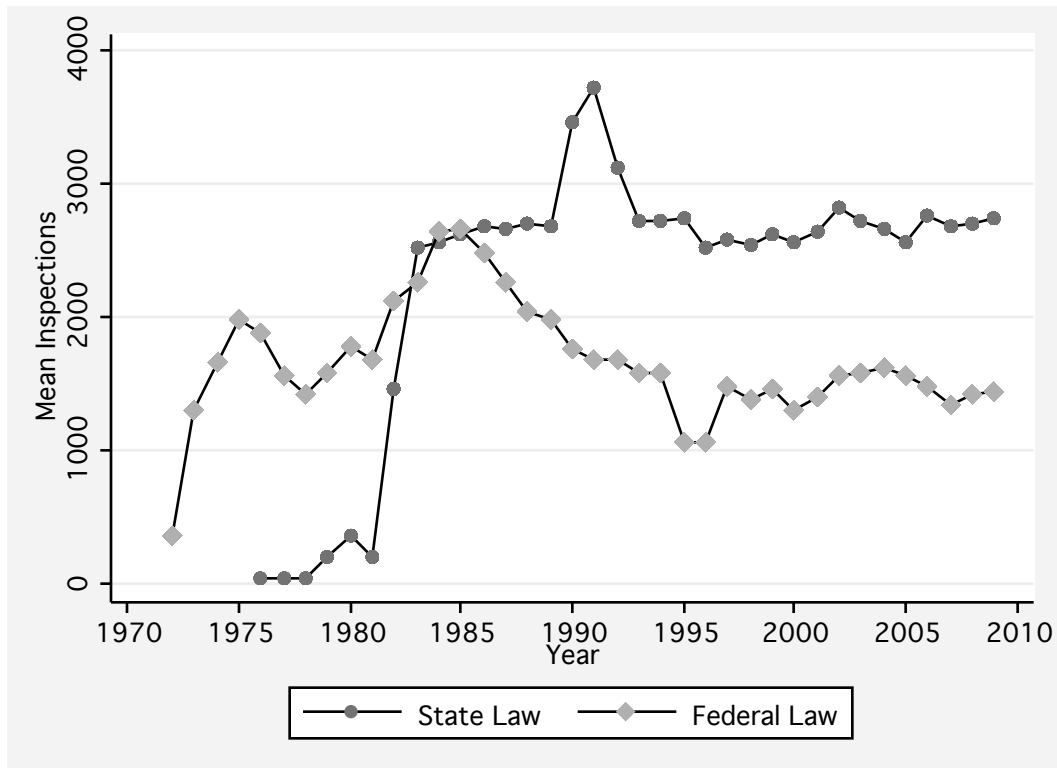


Figure 4.5. Trends in Mean Inspections by Regulatory Status

Note: Data from IMIS.

There is a clear accelerated increase in inspections and violations after certification, one which persists for at least 10 years after certification, and not much of an upward trend prior to certification.

Further, Figure 4b and 5b show trends in mean inspections and violations for states with state regulated OSHA programs, compared to those that are federally regulated. Federal inspections and violations over time are much smoother than state inspections and violations, generally reflective of a fairly constant trend over time. Thus, the finding of a positive coefficient on state OSHA certification is not reflective of these states cutting back inspections and violations less than federally regulated states. Instead, it appears there was an actual increase in inspections and violations after certification.

An increase in inspections might reflect the hiring of more inspectors, or a greater frequency of inspections. An increase in violations might be a natural consequence of increased inspections. On the other hand, an increase in violations might reflect the ability of state inspectors to better target unsafe local businesses and firms.

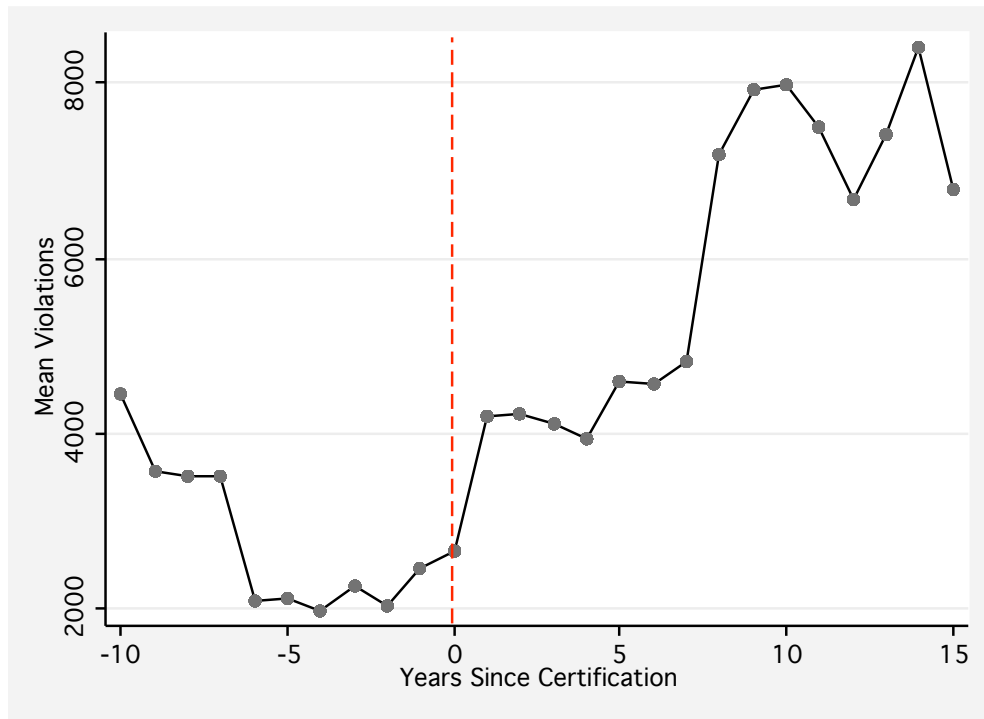


Figure 4.6. Trends in Mean Violations for State Regulated OSHA Programs

Note: Data from IMIS.

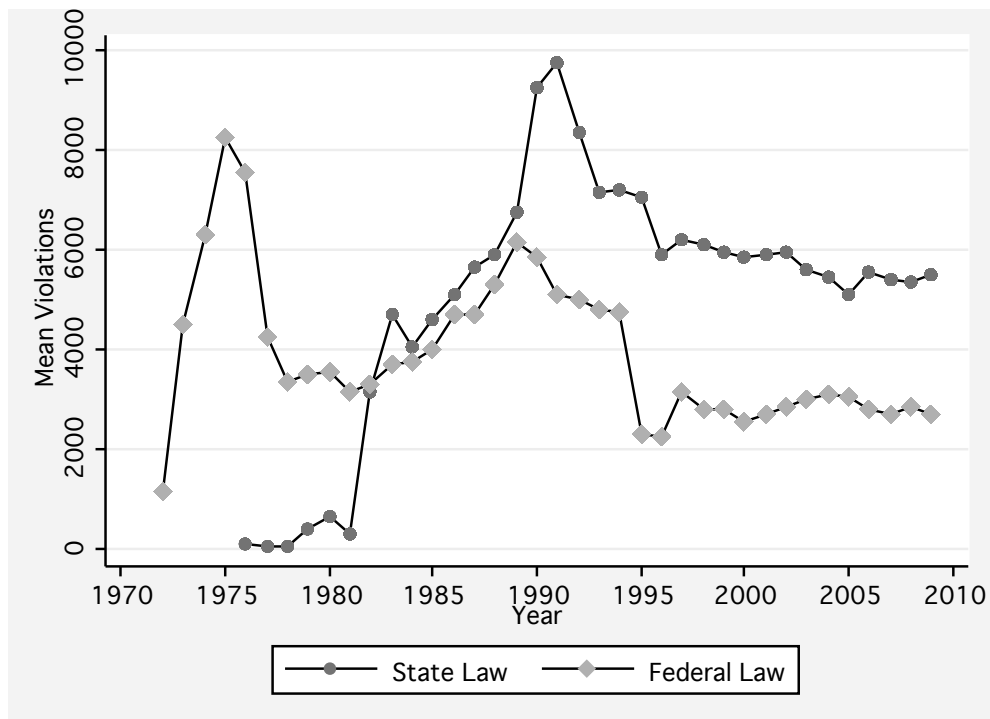


Figure 4.7. Trends in Mean Violations by Regulatory Status

Note: Data from IMIS.

Furthermore, one can break down the number of inspections and violations by industry division. While there appears to be an increase in inspections and violations in all industries following certification, the largest increases are in the construction and manufacturing divisions, which are the most dangerous. See Figures 4.8 and 4.9. The clear break in the trends in both mean inspections and violations in these dangerous divisions further confirms that state OSHA certification was effective at increasing the use of traditional enforcement tools, particularly in the most dangerous industries.

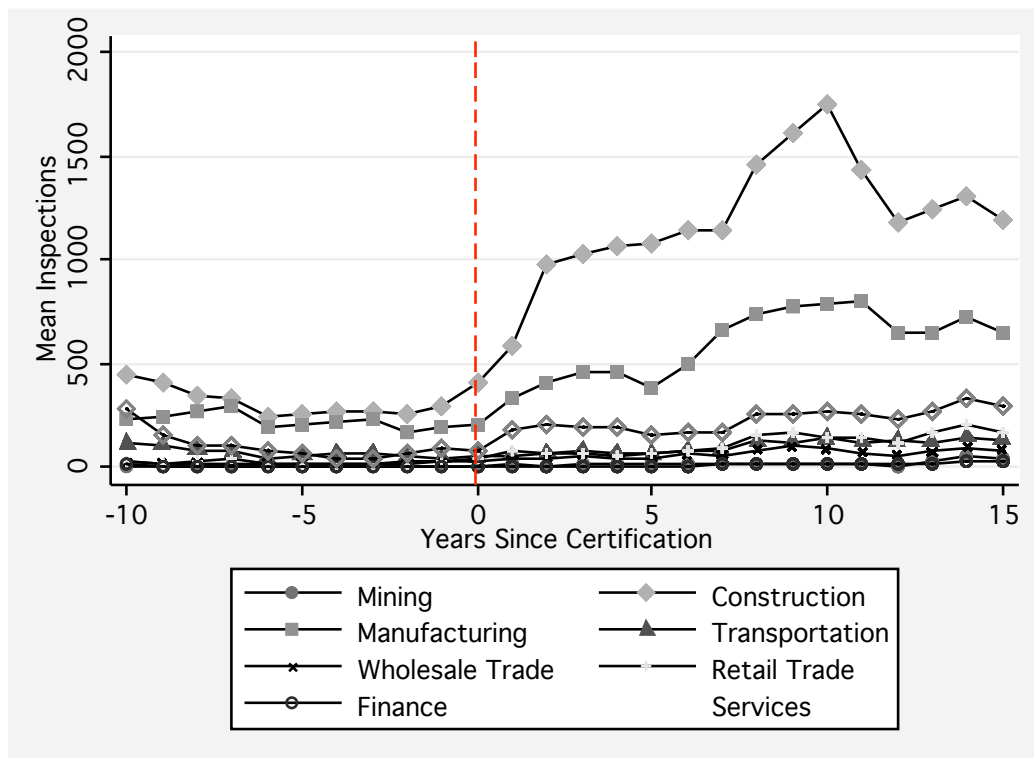


Figure 4.8. Trends in Mean Inspections, by Industry Division

Note: Data from IMIS.

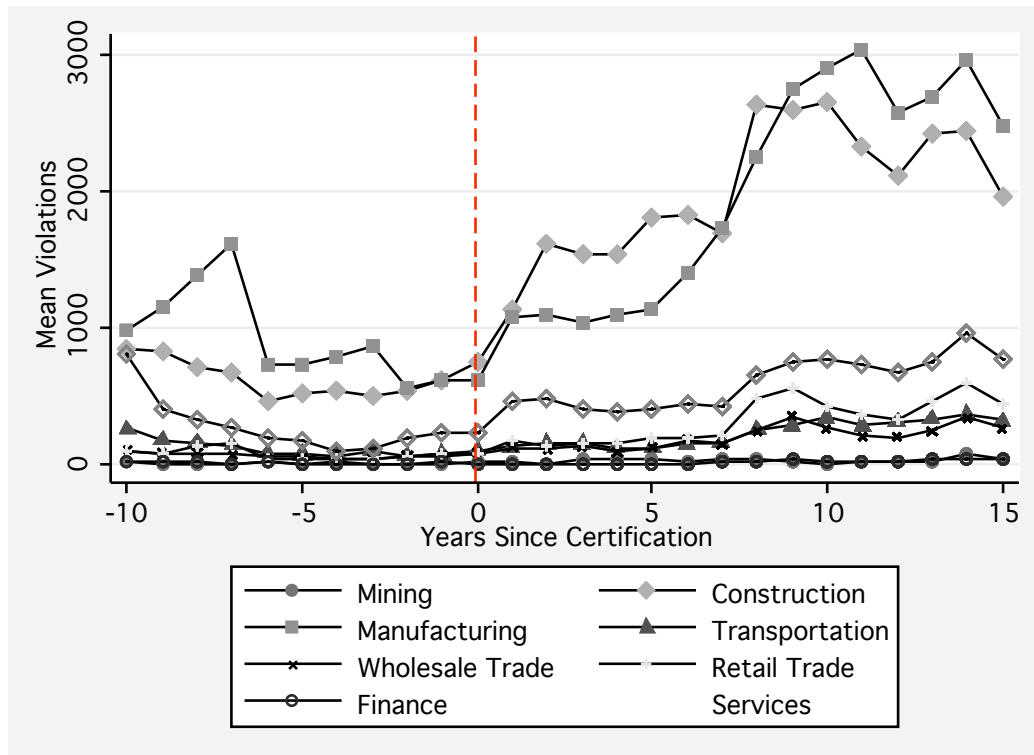


Figure 4.9. Trends in Mean Violations, by Industry Division

Note: Data from IMIS.

It is also interesting to note that certification of state OSHA programs led to an almost universal increase in all types of inspections. Inspections that were due to followups and inspections that were planned increased significantly following certification. The number of state level inspections increased significantly, but no increase in federal level inspections. This is expected since federal enforcement is essentially suspended at this point in time. Complete and partial inspections, which describe the scope of the inspection, also increased following state certification. See Table 4.7.

Table 4.7. Impact of State OSHA Law Certification on Types of Inspections

	(1) Accidents	(2) Follow	(3) Planned	(4) State	(5) Federal	(6) Complete	(7) Partial
State Law	-3.417 (2.25)	7.016* (3.80)	114.0*** (22.7)	132.5*** (25.6)	-11.17 (21.6)	109.0*** (20.8)	18.28* (9.65)
Complaints	0.249*** (0.0037)	0.226*** (0.0062)	0.992*** (0.037)	1.871*** (0.042)	0.966*** (0.035)	1.254*** (0.034)	1.288*** (0.016)
Ind. % Union	-17.30 (14.7)	-4.572 (24.8)	-184.7 (148)	-365.4** (167)	125.3 (141)	-195.4 (136)	-74.29 (63.0)
Ind. Size	-0.00703*** (0.00078)	0.0186*** (0.0013)	-0.0544*** (0.0079)	-0.0349*** (0.0088)	-0.0416*** (0.0075)	-0.0857*** (0.0072)	0.0173*** (0.0033)
Ind. Average Age	-0.631*** (0.13)	0.983*** (0.22)	0.458 (1.29)	-5.782*** (1.45)	3.999*** (1.23)	-0.178 (1.19)	-2.596*** (0.55)
Observations	11552	11552	11552	11552	11552	11552	11552
R^2	0.56	0.32	0.46	0.45	0.46	0.50	0.65

Note: The dependent variable in all regressions is a particular type of inspection. Data are industry-state-year observations. State, industry and year fixed effects are included in all specifications. These regressions include only industry level controls, so data is from 1979-2009. Coefficients that are significant at the .1, .05, .01 percent level are indicated with *, **, ***, respectively.

4.6.1. Differential Impact of State OSHA Programs

State OSHA programs are likely to differ in their original purpose and in the degree to which they adhere to federal standards. From the results in Table 4.8, it appears that certification of an OSHA program initially created under a Democratic governor is associated with a significant increase in the number of inspections per capita, and violations per capita, compared to federally regulated states. At the same time, certification of an OSHA program created under a Republican governor also leads to a significant increase in inspections and violations per capita, and also the violation rate, compared to federally regulated states. One possible hypothesis is that more conservative states sought to avoid federal enforcement and regulation by the creation of their own state OSHA programs. While inspections and violations may have increased, this does not indicate that workplace injury rates were reduced. This will be explored in the next section.

Additionally, it is often the case that the party affiliation of the governor at the time of state OSHA creation differs from the overall political climate that followed. If states with OSHA programs initially created under a Republican governor eventually became more liberal, one might see changes in inspections and violations over time. There have been very few changes and amendments to the original state OSHA laws over time, but enforcement could have changed as the politics changed within states.

Table 4.8. Impact of State OSHA Laws on Inspections/Capita, Violations/Capita and Violation Rate, By Political Climate

	(1) Inspections	(2) Violations	(3) Viol Rate	(4) Inspections	(5) Violations	(6) Viol Rate
Certified - Republican	0.160*** (0.037)	0.367*** (0.074)	0.107** (0.048)			
Certified - Democrat				0.0637*** (0.019)	0.123*** (0.037)	-0.0233 (0.040)
Employee Complaints	0.000231* (0.00012)	0.000733* (0.00039)	0.00000839 (0.000055)	0.0000850* (0.000047)	0.000237* (0.00013)	0.0000136 (0.000026)
Industry % Union	-0.0935 (0.15)	-0.388 (0.33)	-0.0202 (0.21)	-0.0695 (0.051)	-0.0817 (0.064)	-0.180 (0.17)
Industry Size	-0.00000429 (0.000011)	-0.0000237 (0.000027)	0.00000596 (0.0000065)	0.00000984* (0.0000052)	0.0000102 (0.000011)	-0.00000105 (0.0000055)
Industry Average Age	-0.0000745 (0.0012)	-0.00179 (0.0025)	0.000535 (0.0012)	0.000298 (0.00097)	0.0000946 (0.0017)	-0.00206 (0.0015)
Observations	6559	6559	8156	7444	7444	9284
R^2	0.56	0.52	0.21	0.50	0.51	0.21

Note: The dependent variable in all regressions is either total number of inspections, violations or violation rate. Data are industry-state-year observations. State, industry and year fixed effects are included in all specifications. Standard errors are clustered at the state level. These regressions include only industry level controls, so data is from 1979-2009. Coefficients that are significant at the .1, .05, .01 percent level are indicated with *, **, ***, respectively.

Figures 4.10 and 4.11 show that indeed, state OSHA programs implemented under both Republican and Democratic governors experienced an increase in mean inspections and violations in the years following certification, although the relative increase was larger for states with Republican governors.

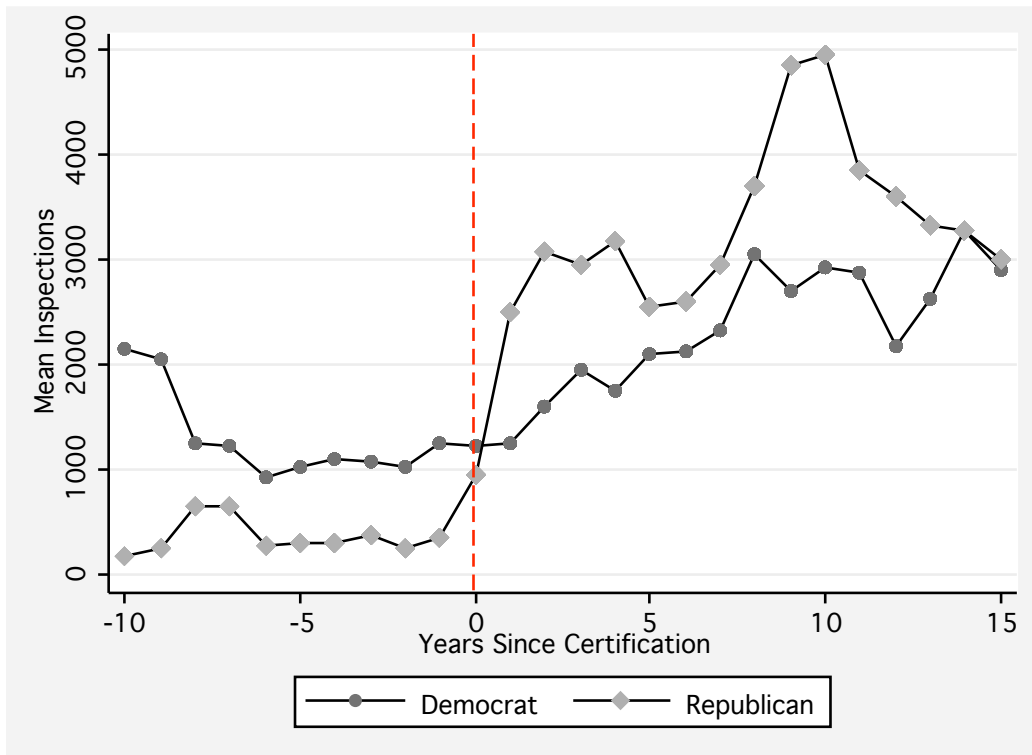


Figure 4.10. Trends in Mean Inspections, By Political Climate

Note: Data from IMIS.

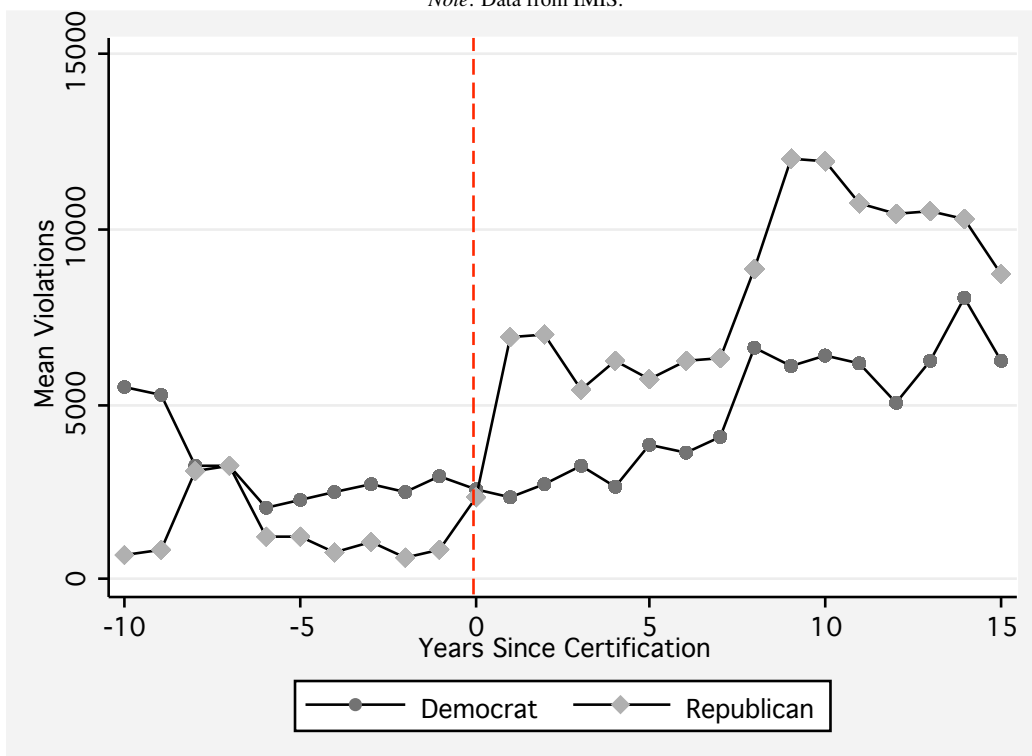


Figure 4.11. Trends in Mean Violations, By Political Climate

Note: Data from IMIS.

In addition to state OSHA programs being implemented for potentially very different purposes, some state programs have adopted rules more rigorous than the federal OSHA, while other state programs are identical to the federal OSHA. Table 4.9 presents the effects of these two types of state OSHA programs on inspections and violations. It appears that state certification of a program with more stringent standards than the federal OSHA lead to a significant increase in inspections per capita, violations per capita, and the violation rate, suggesting a huge rise in the probability of issuing a violation charge. Given that violation charges are viewed as most costly from the perspective of employers, particularly as penalties increased, this would suggest that these states with different standards should also be the most effective at reducing injury rates. This will be explored next.

Table 4.9. Impact of State OSHA Laws on Inspections, Violations and Violation Rate, By Standards

	(1) Inspections	(2) Violations	(3) Viol Rate	(4) Inspections	(5) Violations	(6) Viol Rate
Certified - Diff Standards	0.155*** (0.032)	0.360*** (0.080)	0.137*** (0.044)			
Certified - No Diff Standards				0.0514** (0.022)	0.0865 (0.054)	-0.00777 (0.032)
Employee Complaints	0.000137* (0.000078)	0.000401* (0.00024)	0.0000112 (0.000025)	0.000136** (0.000056)	0.000435** (0.00017)	0.0000210 (0.000045)
Industry % Union	-0.121 (0.084)	-0.270 (0.18)	-0.107 (0.15)	-0.0943 (0.097)	-0.184 (0.21)	-0.197 (0.16)
Industry Size	0.00000346 (0.0000060)	-0.00000741 (0.000014)	-0.00000181 (0.0000051)	0.0000123** (0.0000053)	0.00000643 (0.000014)	-0.00000104 (0.0000071)
Industry Average Age	0.000182 (0.0010)	-0.000983 (0.0020)	-0.000646 (0.0013)	0.000843 (0.00094)	0.000349 (0.0015)	-0.00121 (0.0014)
Observations	9257	9257	11552	8553	8553	10656
R^2	0.50	0.49	0.22	0.50	0.49	0.21

Note: The dependent variable in all regressions is either total number of inspections, violations or violation rate. Data are industry-state-year observations. State, industry and year fixed effects are included in all specifications. These regressions include only industry level controls, so data is from 1979-2009. Standard errors are clustered at the state level. Coefficients that are significant at the .1, .05, .01 percent level are indicated with *, **, ***, respectively.

Figures 4.12 and 4.13 show that indeed, state OSHA programs with different standards had a greater increase in mean inspections and violations in the years following certification, although the increase seems to begin between 5-10 years after certification, potentially due to lags.

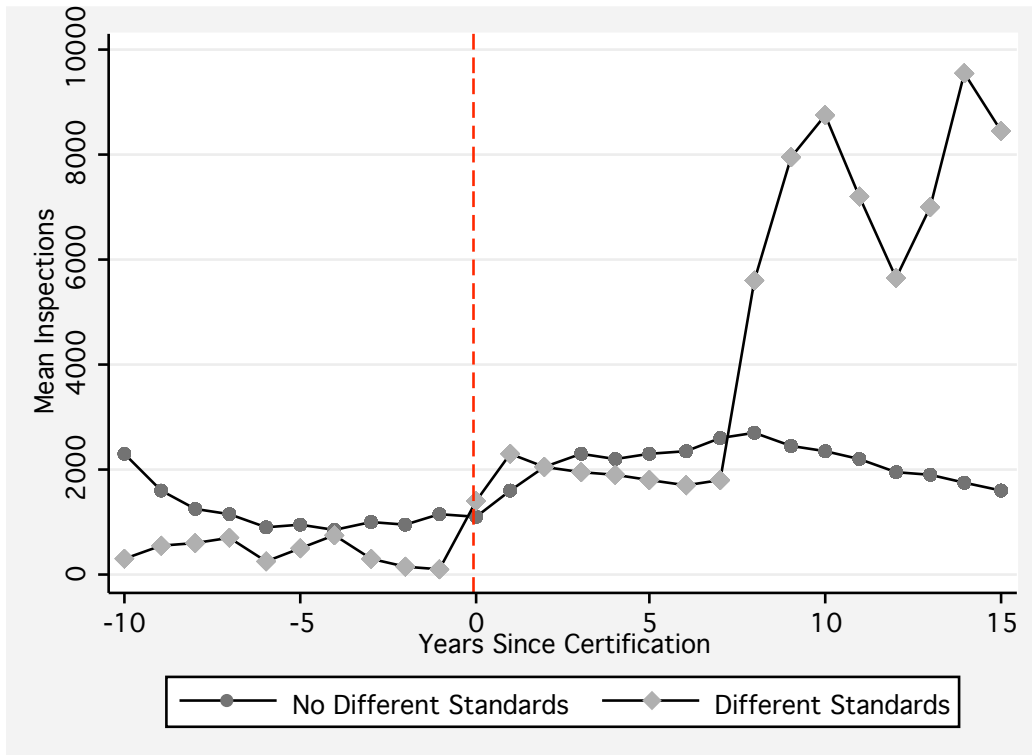


Figure 4.12. Trends in Mean Inspections, By Standards

Note: Data from IMIS.

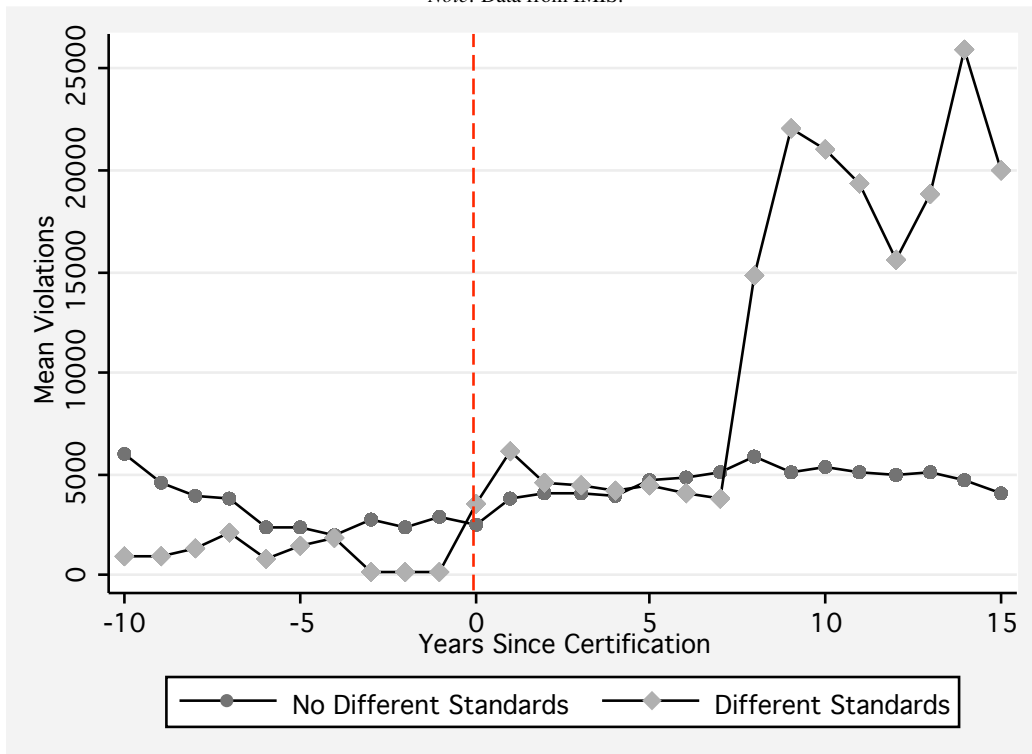


Figure 4.13 Trends in Mean Violations, By Standards

Note: Data from IMIS.

4.7. Results - Impact of State OSHA programs on Injuries

Table 4.10 presents the cross sectional comparison of states with own OSHA programs versus states with federally enforced OSHA programs. The outcome of interest is total nonfatal injuries and illnesses incidence rates as well as the total non fatal injuries and illnesses incidence rates that led to lost workdays. The incidence rates represent the number of injuries and illnesses per 100 full-time workers.

The dummy for state law is the variable of interest. I regress the outcome variables on industry level characteristics and state level workers' compensation variables.

One sees that states with their own OSHA programs are not associated with a significantly lower incidence rate of both total nonfatal injuries and illnesses and those leading to days of lost work from 1996-2008. Nor are states with their own OSHA programs associated with significantly lower number of fatalities, compared to federally regulated states.

Like Morantz, I find that the degree of unionization is not significantly associated with nonfatal injury rates that lead to lost work days, although Weil has argued that unionized workplaces have higher rates of OSHA compliance. Morantz argues that this may be because the link between unionization and nonfatal injuries is less strong. However, I also find an insignificant relationship between unionization and fatalities.

The average age of workers in an industry division does not appear to have a significant effect on total nonfatal injuries, but is positively associated with total nonfatal injuries that lead to lost workdays. It appears that older workers are more likely to be nonfatally injured and take days off from work. However, it seems that the greater the average age, the fewer the number of fatalities, suggesting that younger workers are more likely to be fatally injured. This make sense as younger workers generally select into more dangerous industries.

Table 4.10 also shows that greater coverage of workers provided through the state's workers' compensation programs is associated with a lower rate of nonfatal injuries, but a higher rate of fatalities. This may indicate that workers' compensation coverage is more expansive in states with more dangerous industry composition. Furthermore, total workers' compensation benefits are positively associated with nonfatal injuries leading to lost work days. The positive relationship between the workers' compensation replacement rate, benefits as a proportion of payroll, and lost work nonfatal injuries potentially indicates a moral hazard effect, whereby workers are more likely to report being injured the greater their wage replacement. Similarly, the positive relationship might indicate that workers are just more likely to claim (which takes time and effort), the greater the replacement rate.

Table 4.10. Impact of State OSHA Laws on Non Fatal Injuries and Illnesses 1996-2008, Fatal Injuries 2003-2008

	(1) NonFatal	(2) Lost Work	(3) Fatal
Dummy for State Law	-0.300 (0.64)	0.172 (0.13)	-0.412 (0.96)
Industry % Union	2.684 (2.35)	0.171 (1.33)	3.014 (15.3)
Industry Size	-0.000259 (0.00019)	-0.00000335 (0.000057)	0.0000451 (0.00078)
Industry Average Age	0.0332 (0.026)	0.0730*** (0.013)	-0.463** (0.19)
WC Covered Workers	-0.000273*** (0.000095)	-0.000104** (0.000046)	0.00415*** (0.0011)
Total WC Benefits	0.000000361** (0.00000015)	0.000000138** (0.000000064)	-0.00000306* (0.0000016)
WC % Rate	0.341 (0.38)	0.344*** (0.094)	1.613 (1.25)
Observations	5042	5031	2100
R^2	0.05	0.31	0.41

Note: The dependent variable in regressions (1) and (2) is the total nonfatal injury and illness incidence rates and the incidence rates for cases leading to total lost workdays. Incidence rates represent the number of injuries and illnesses per 100 full time workers. The dependent variable in regressions (3) is the total number of fatalities. These data are from the Survey of Occupational Injuries and Illnesses, from 1996-2008, and the Census of Fatal Occupational Injuries, from 2003-2008. Data are industry-state-year observations. Industry and year fixed effects are included in all specifications. Standard errors are clustered at the state level. Coefficients that are significant at the .1, .05, .01 percent level are indicated with *, **, ***, respectively.

4.7.1. Differential Impact of State OSHA Programs

It is also interesting when one disaggregates state regulated OSHA programs by political climate at time of creation of state OSHA program and by their standards, as compared to all other states. Table 4.11 shows that state OSHA programs created by a Republican governor have been ineffective at reducing the incidence rate of total nonfatal injuries and those resulting in lost workdays from 1996-2008 at the industry level, compared to all other states. Additionally, state OSHA programs implemented by a Republican were ineffective at reducing fatalities from 2003-2008. Similarly, state programs created by Democratic governors appear to be ineffective at reducing both nonfatal injury rates and the total number of fatalities.

As seen in Table 4.8, state program certification was associated with an increase in the use of traditional enforcement tools such as inspections per capita and violations per capita, particularly among state programs created by a Republican governor. However, data on injuries and fatalities in the more present period suggest that there are no significant differences between state regulated OSHA programs (both Democratic and Republican), compared to federally regulated OSHA states.

Table 4.11. Impact of State OSHA Laws on Non Fatal Injuries and Illnesses, 1996-2008 and Fatalities, 2003-2008, By Political Climate

	(1) NonFatal	(2) Lost Work	(3) Fatal	(4) NonFatal	(5) Lost Work	(6) Fatal
Dummy for State Law - Republican	-0.368 (0.69)	0.0791 (0.18)	-0.609 (1.01)			
Dummy for State Law - Democrat				-0.437 (0.69)	0.201 (0.13)	-1.545 (1.14)
Industry % Union	5.120 (3.38)	1.216 (1.83)	-14.90 (15.5)	1.461 (2.70)	-0.454 (1.51)	17.84 (24.4)
Industry Size	-0.000322 (0.00031)	-0.000000631 (0.000091)	-0.00113 (0.0010)	-0.000318 (0.00023)	-0.0000104 (0.000057)	-0.0000292 (0.00083)
Industry Average Age	0.0399 (0.030)	0.0656*** (0.014)	-0.376** (0.17)	0.0279 (0.040)	0.0854*** (0.015)	-0.522** (0.24)
WC Covered Workers	-0.000120 (0.00018)	-0.0000597 (0.000054)	0.00622*** (0.00056)	-0.000249** (0.000094)	-0.0000879** (0.000043)	0.00412*** (0.0012)
WC Benefits	-0.000000137 (0.00000048)	2.55e-09 (0.00000013)	-0.00000811*** (0.0000015)	0.000000351** (0.00000015)	0.000000115* (0.000000057)	-0.00000289* (0.0000016)
WC Rep Rate	0.335 (0.40)	0.338*** (0.095)	3.342* (1.65)	0.368 (0.41)	0.365*** (0.10)	1.141 (1.15)
Observations	3607	3598	1561	4054	4046	1748
R ²	0.04	0.39	0.43	0.04	0.28	0.41

Note: The dependent variable in regressions (1) and (4) is the total nonfatal injury and illness incidence rates and in regressions (2) and (5) the incidence rates for cases leading to total lost workdays. Incidence rates represent the number of injuries and illnesses per 100 full time workers. The dependent variable in regressions (3) and (6) is the total number of fatalities. These data are from the Survey of Occupational Injuries and Illnesses, from 1996-2008, and the Census of Fatal Occupational Injuries, from 2003-2008. Data are industry-state-year observations. Industry and year fixed effects are included in all specifications. Standard errors are clustered at the state level. Coefficients that are significant at the .1, .05, .01 percent level are indicated with *, **, ***, respectively.

When I disaggregate state OSHA programs into those with different standards from the federal program and those identical to the federal program, I find that state OSHA programs that are more stringent than the federal program are associated with no significant differences in total nonfatal injury rates, but seem to have a higher proportion of nonfatalities leading to lost workdays. See Table 4.12.

Table 4.12 also shows that state OSHA programs with different standards have a substantially lower number of fatalities than other programs, whereas state OSHA programs identical to federal standards have no significant difference in fatalities, from 2003-2008. Compared to the mean industry number of fatalities from 2003-2008 of 27.45, this reduction by 2.799 is an approximately 10% reduction, suggestive that more stringent OSHA standards reduce workplace fatalities.

Overall, the results suggest that states with different standards for their own OSHA programs also issue more inspections per capita, violations per capita and have a higher violation rate, as seen in Table 4.9. This increased use of enforcement tools can potentially explain the significant reduction of fatalities in these states.

The effect of federal versus state enforcement is most clear when one just compares state regulated OSHA programs identical to federal standards to federally regulated OSHA programs. Table 4.9 shows that when states adopt programs identical to the federal program, there is a significant increase in inspections per capita. However, Table 4.12 presents evidence that there is no reduction in total nonfatal injury rates and fatalities from 1996-2008.

Table 4.12. Impact of State OSHA Laws on Non Fatal Injuries and Illnesses, 1996-2008 and Fatalities, 2003-2008, By Standards

	(1) NonFatal	(2) Lost Work	(3) Fatal	(4) NonFatal	(5) Lost Work	(6) Fatal
Dummy for State Law - Different	0.585 (0.44)	0.467*** (0.084)	-2.799* (1.66)			
Dummy for State Law - Not Different				-0.521 (0.66)	0.0882 (0.15)	-0.232 (0.78)
Industry % Union	2.822 (2.39)	0.228 (1.33)	2.403 (14.8)	2.423 (3.17)	0.364 (1.77)	-8.161 (18.5)
Industry Size	-0.000228 (0.00016)	-0.00000656 (0.000057)	0.00000179 (0.00078)	-0.000291 (0.00029)	0.0000332 (0.000091)	-0.00119 (0.00096)
Industry Average Age	0.0421** (0.019)	0.0705*** (0.012)	-0.443** (0.19)	0.0333 (0.029)	0.0700*** (0.012)	-0.282** (0.13)
WC Covered Workers	-0.000218** (0.000084)	-0.000102** (0.000039)	0.00407*** (0.0011)	-0.000114 (0.00016)	-0.0000820 (0.000061)	0.00621*** (0.00054)
WC Benefits	0.000000208* (0.00000011)	0.0000000970* (0.000000050)	-0.00000273* (0.0000016)	-0.000000213 (0.00000041)	0.0000000265 (0.00000015)	-0.00000796*** (0.0000015)
WC Rep Rate	0.406 (0.33)	0.316*** (0.083)	1.698 (1.22)	0.383 (0.39)	0.335*** (0.092)	3.399** (1.55)
Observations	5042	5031	2100	4533	4521	1878
R^2	0.05	0.31	0.42	0.05	0.31	0.43

Note: The dependent variable in regressions (1) and (4) is the total nonfatal injury and illness incidence rates and in regressions (2) and (5) the incidence rates for cases leading to total lost workdays. Incidence rates represent the number of injuries and illnesses per 100 full time workers. The dependent variable in regressions (3) and (6) is the total number of fatalities. These data are from the Survey of Occupational Injuries and Illnesses, from 1996-2008, and the Census of Fatal Occupational Injuries, from 2003-2008. Data are industry-state-year observations. Industry and year fixed effects are included in all specifications. Standard errors are clustered at the state level. Coefficients that are significant at the .1, .05, .01 percent level are indicated with *, **, ***, respectively.

4.8. 2000 California Penalty Change

The following tables show the effects of the California penalty increases on issuance of inspections and violations, and also the ultimate policy outcomes of interest, nonfatal and fatal injuries.

The dependent variable of interest, Penalty Change, equals 1 in California, in each year after the penalty change was implemented. In each of the tables, there are three sets of control states that I use and results are robust to choice of control group. In the first specification, I use all other states as controls to California. In the second specification, I include only those other states with state regulated OSHA programs as controls. Finally, in the third specification, I include only the three other states, with state regulated OSHA programs, and which have more restrictive standards than the federal program.

Table 4.13 shows that the penalty change in California was associated with a significant increase in both the number of inspections and the number of violations. Inspections in California increased compared to all other states, all other states with state regulated OSHA, and all other state regulated OSHA program with more stringent standards than the federal program. Violations in California increased compared to all other states and compared to other states with state regulated OSHA programs. However, the number of violations between California and the other three states with more stringent OSHA programs did not differ significantly following the penalty change.

Table 4.13. CA Penalty Change on Inspections, Violations and Violation Rate

	(1) Inspections	(2) Inspections	(3) Inspections	(4) Violations	(5) Violations	(6) Violations	(7) Viol Rate	(8) Viol Rate	(9) Viol Rate
Penalty Change	364.4*** (47.8)	379.7*** (57.6)	204.8** (96.8)	705.0*** (110)	753.0*** (142)	439.2 (270)	0.0150 (0.027)	0.0198 (0.025)	0.0202 (0.025)
Employee Complaints	2.849*** (0.042)	2.783*** (0.059)	1.655*** (0.12)	7.313*** (0.098)	6.769*** (0.15)	4.052*** (0.33)	0.0000188 (0.000024)	0.0000140 (0.000026)	-0.0000203 (0.000030)
Industry % Union	-201.5 (170)	-296.0 (250)	-765.3 (718)	-551.2 (391)	-590.8 (616)	-2091 (2008)	-0.103 (0.095)	-0.0491 (0.11)	0.315* (0.18)
Industry Size	-0.0752*** (0.0090)	-0.0494*** (0.013)	-0.0460 (0.031)	-0.0307 (0.021)	0.0499 (0.033)	-0.00661 (0.087)	-0.00000218 (0.0000050)	-0.00000849 (0.0000058)	0.00000288 (0.0000080)
Industry Average Age	-1.764 (1.48)	-0.860 (2.51)	-3.042 (8.38)	3.592 (3.40)	3.849 (6.19)	7.639 (23.4)	-0.000680 (0.00082)	-0.000371 (0.0011)	0.00437** (0.0021)
Observations	11552	5438	896	11552	5438	896	11552	5438	896
R^2	0.65	0.68	0.72	0.65	0.63	0.64	0.22	0.26	0.35

Note:

The dependent variable in all regressions is either total number of inspections, violations or violation rate. Regressions (1), (4) and (7) include all other states as controls. Regressions (2), (5), and (8) include states with own OSHA programs as controls. Regressions (3), (6) and (9) include only other state OSHA programs with different standards as controls. Data are industry-state-year observations. State, industry and year fixed effects are included in all specifications. These regressions include only industry level controls, so data is from 1979-2009. Coefficients that are significant at the .1, .05, .01 percent level are indicated with *, **, ***, respectively.

Table 4.14 shows that subsequently, there was no significant change in the incidence rate for nonfatal injuries at the industry level from 1996-2008. While the penalty change does not seem to be associated with reductions in nonfatalities, there is a significant decrease in the number of workplace fatalities at the aggregate private industry level during the same time period. See Table 4.15. When compared to all other states, the penalty change is associated with a 113.3 decrease in fatalities. The mean number of private industry fatalities in all states during this time was 111.0876, so the penalty reform corresponded with an over 100% reduction in fatalities. When compared to the control group of states with state regulated OSHA programs, the penalty reform leads to a 84.12 reduction in fatalities. Compared to the mean number of private industry fatalities in states with state regulated OSHA programs of 118.1632, the penalty reform was associated with a 70% decrease in fatalities. However, compared to the three other states with more stringent standards than the federal program, with mean fatalities of 205.96, the penalty reform does not appear to have been any more effective at reducing fatalities.

Table 4.14. CA Penalty Change on NonFatal Injuries, 1996-2008

	(1) NonFatal	(2) NonFatal	(3) NonFatal	(4) Lost Work	(5) Lost Work	(6)) Lost Work
Penalty Change	0.0145 (0.86)	0.0847 (0.59)	-0.400 (0.50)	-0.0140 (0.28)	-0.124 (0.52)	0.0509 (0.53)
Industry % Union	3.905 (3.27)	0.369 (1.49)	1.762 (0.78)	0.0487 (1.21)	-0.931 (1.24)	-0.799 (0.77)
Industry Size	-0.000374** (0.00015)	-0.000184** (0.000082)	-0.0000847** (0.000026)	-0.000111** (0.000053)	-0.0000589 (0.000071)	0.0000181 (0.000044)
Industry Average Age	0.0475** (0.018)	0.0550** (0.020)	0.0601** (0.018)	0.0669*** (0.012)	0.0701*** (0.019)	0.101 (0.071)
WC Covered Workers	-0.0000811 (0.00023)	0.00000913 (0.00033)	0.000530 (0.00040)	0.0000489 (0.000031)	0.0000657 (0.00023)	0.000461 (0.00035)
Total WC Benefits	0.0000000948 (0.00000031)	0.000000171 (0.00000011)	0.000000218* (0.000000089)	0.0000000836 (0.000000091)	0.000000119 (0.00000012)	-0.000000123 (0.000000088)
WC Rep Rate	-0.222 (0.51)	-0.255 (0.44)	-0.470 (0.29)	0.136 (0.22)	-0.103 (0.28)	0.556* (0.19)
Observations	5042	2423	509	5031	2418	510
R^2	0.11	0.54	0.59	0.35	0.30	0.35

Note: The dependent variable in regressions (1) - (3) is the total nonfatal injury and illness incidence rates and in regressions (4) -(6) the incidence rates for cases leading to total lost workdays. Incidence rates represent the number of injuries and illnesses per 100 full time workers. Data for nonfatalities are from the Survey of Occupational Injuries and Illnesses, 1996-2008. Regressions (1), (4) include all other states as controls. Regressions (2), (5), and (8) include states with own OSHA programs as controls. Regressions (3), (6) include only other state OSHA programs with different standards as controls. Data are state-year observations for the private industry. State, industry, and year fixed effects are included in all specifications. Standard errors are clustered at the state level. Coefficients that are significant at the .1, .05, .01 percent level are indicated with *, **, ***, respectively.

Table 4.15. CA Penalty Change on Fatalities, 1996-2008
Private Industry Totals

	(1) Fatal	(2) Fatal	(3) Fatal
Penalty Change	-113.3*** (18.9)	-84.12*** (28.6)	28.29 (53.6)
Industry % Union	668.7*** (242)	63.43 (282)	7639* (2734)
Industry Size	0.00385 (0.0024)	0.00170 (0.0058)	0.0185* (0.0069)
Industry Average Age	-0.528 (1.61)	-2.683 (2.96)	3.267 (19.1)
WC Covered Workers	-0.00230 (0.0029)	-0.00743 (0.024)	-0.0817** (0.021)
Total WC Benefits	-0.0000158** (0.0000063)	-0.0000247*** (0.0000050)	-0.0000330* (0.000011)
WC Rep Rate	2.300 (6.69)	6.424 (6.16)	82.73 (36.5)
Observations	672	246	44
R^2	0.98	0.98	0.99

Note: The dependent variable in the regressions is the total number of fatalities. 1996-2008. Data for fatalities are from the Census of Fatal Occupational Injuries, from 1996-2008. Regression (1) includes all other states as controls, regression (2) include states with own OSHA programs as controls, and regression (3) includes only other state OSHA programs with different standards as controls. Data are state-year observations for the private industry. State and year fixed effects are included in all specifications. Standard errors are clustered at the state level. Coefficients that are significant at the .1, .05, .01 percent level are indicated with *, **, ***, respectively.

Figure 4.14 shows that after the California penalty reform in 2000, there was no change in the nonfatal injury incidence rate in California beyond the trend of a decreasing nonfatal injuries. However, Figure 4.15 shows that total fatalities did appear to decrease post 2000 reform in California, while the trend in mean fatalities in all other states was relatively constant.

Ultimately, it seems that increases in civil penalty maximums and criminal penalties, and thus the magnitude and type of sanction, can be an effective way to increase workplace safety. Increased penalties, particularly criminal sanctions, lead to an increase in traditional enforcement tools in the form of issuance of inspections and violations, and reductions in the number of fatalities.

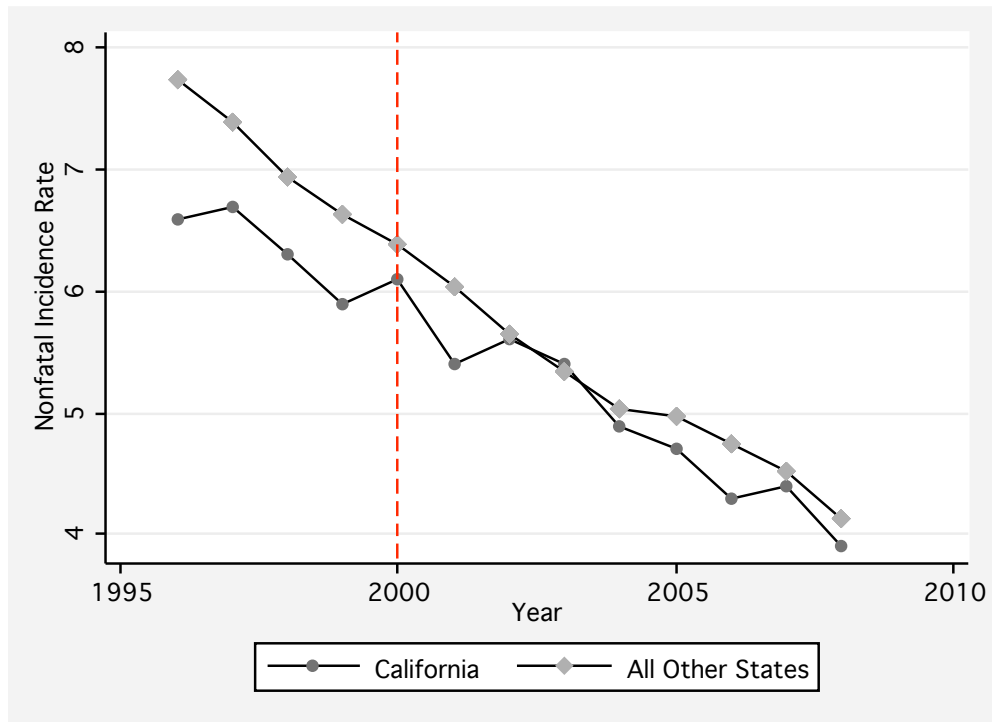


Figure 4.14. Trends in Nonfatal Injury Rates

Note: Data from SOII, 1996-2008.

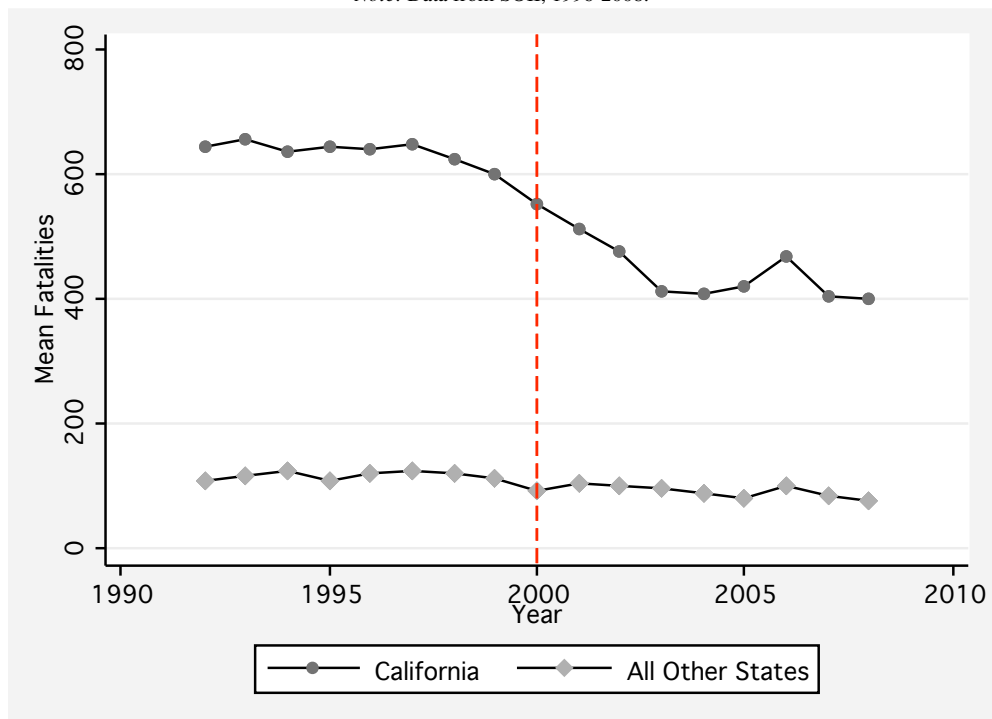


Figure 4.15. Trends in Mean Fatalities

Note: Data from CFOI, 1992-2008.

4.9. Results - Impact of State OSHA programs on Wages and Employment

Table 4.16 presents the results of the impact of a state regulated OSHA certification, on the log real wages of workers. Specification 1 shows the results for all employed workers. Specification 2 looks only at workers in the private sector, who were covered by federal OSHA before a state regulated OSHA program was certified. Finally, specification 3 presents the results for those workers in the public sector, who had no mandated workplace safety regulation prior to the creation of state regulated OSHA programs.

	(1) All Workers	(2) Private Sector	(3) Public Sector
State Law/Certified	-0.00594*** (0.0015)	-0.0201*** (0.0020)	0.00626* (0.0036)
Age	0.00644*** (0.000017)	0.00607*** (0.000020)	0.00794*** (0.000044)
Female	-0.207*** (0.00045)	-0.215*** (0.00053)	-0.160*** (0.0011)
Black	-0.0979*** (0.00078)	-0.111*** (0.00094)	-0.0600*** (0.0018)
Hispanic	-0.124*** (0.00086)	-0.140*** (0.00099)	-0.0421*** (0.0026)
Other - NonWhite	-0.0807*** (0.0011)	-0.0915*** (0.0013)	-0.0391*** (0.0026)
HS Degree	0.196*** (0.00068)	0.190*** (0.00075)	0.224*** (0.0023)
Some College	0.307*** (0.00072)	0.301*** (0.00080)	0.319*** (0.0023)
College Degree	0.622*** (0.00080)	0.613*** (0.00093)	0.628*** (0.0023)
Advanced Degree	0.812*** (0.0010)	0.799*** (0.0013)	0.837*** (0.0024)
Married	0.138*** (0.00046)	0.144*** (0.00053)	0.101*** (0.0012)
Union Coverage	0.106*** (0.0015)	0.103*** (0.0023)	0.104*** (0.0020)
Observations	4678018	3464320	654721
R ²	0.42	0.42	0.40

Note: The dependent variable in all regressions is the log real wage, in 2009 dollars. Data are individual level observations, from the 1979-2009 CPS ORG. State, 2 digit industry and year fixed effects are included in all specifications. Coefficients that are significant at the .1, .05, .01 percent level are indicated with *, **, ***, respectively.

Overall, it appears that the certification of a state regulated OSHA program led to a .594% decrease in real wages. This fall in wages seems to be driven by the decrease in wages of private sector workers following certification, by about 2%. There does not seem to be a significant change in the wages of public sector workers who were not covered under any federal workplace safety mandate before the certification of state regulated OSHA programs. It could be however, that public sector industries had already taken workplace safety measures.

Table 4.17 presents the wage results broken down by industries within the private sector. The results suggest that the compensating differential in wages is largest for those industries that are the most dangerous, in particular, the most dangerous, construction. Wages in the construction industry fall by 1.5% after certification of a state regulated OSHA program. There are also falls in manufacturing, wholesale trade, retail trade and services.

Table 4.17. Impact of State OSHA Laws on Log Real Wages, by Private Sector

	(1) Agriculture	(2) Mining	(3) Construction	(4) Manufacturing	(5) Transp	(6) Whole. Trade	(7) Ret. Trade	(8) Finance	(9) Services
State Law/Certified	0.0485*** (0.014)	-0.00762 (0.015)	-0.0151** (0.0072)	-0.0474*** (0.0038)	-0.00730 (0.0080)	-0.0206** (0.0097)	-0.0168*** (0.0043)	-0.00841 (0.0084)	-0.0175*** (0.0044)
Age	0.00529*** (0.00011)	0.00786*** (0.00019)	0.00853*** (0.000080)	0.00764*** (0.000038)	0.00802*** (0.00010)	0.00725*** (0.000096)	0.00642*** (0.000044)	0.00644*** (0.000084)	0.00508*** (0.000037)
Female	-0.153*** (0.0029)	-0.250*** (0.0061)	-0.246*** (0.0029)	-0.253*** (0.00096)	-0.208*** (0.0024)	-0.238*** (0.0025)	-0.256*** (0.0011)	-0.236*** (0.0022)	-0.160*** (0.0010)
Black	-0.0182*** (0.0016)	-0.0437*** (0.0033)	-0.0754*** (0.0013)	-0.0618*** (0.00062)	-0.0503*** (0.0017)	-0.0670*** (0.0017)	-0.0361*** (0.00081)	-0.0583*** (0.0016)	-0.0469*** (0.00065)
Education	0.119*** (0.0014)	0.179*** (0.0021)	0.133*** (0.0010)	0.223*** (0.00045)	0.167*** (0.0012)	0.211*** (0.0012)	0.131*** (0.00060)	0.199*** (0.0011)	0.236*** (0.00043)
Married	0.129*** (0.0031)	0.116*** (0.0052)	0.148*** (0.0020)	0.139*** (0.0010)	0.144*** (0.0025)	0.138*** (0.0025)	0.173*** (0.0012)	0.112*** (0.0022)	0.160*** (0.0010)
Union Coverage	0.136*** (0.024)	0.0505*** (0.016)	0.163*** (0.0084)	0.0811*** (0.0038)	0.139*** (0.0063)	0.0555*** (0.014)	0.147*** (0.0070)	0.0162 (0.013)	0.0827*** (0.0045)
Observations	129053	38578	204211	866317	168394	160887	585137	220945	1044056
R^2	0.17	0.30	0.27	0.40	0.27	0.33	0.32	0.29	0.36

Note: The dependent variable in all regressions is the log real wage, in 2009 dollars. Data are individual level observations, from the 1979-2009 CPS ORG. State, 2 digit industry and year fixed effects are included in all specifications. Coefficients that are significant at the .1, .05, .01 percent level are indicated with *, **, ***, respectively.

While the wages of all workers fell following state law certification, the mean number of hours increased significantly by 0.0615. For private sector workers, however, the fall in wages was accompanied by no change in hours worked, while for public sector workers, there was a significant fall in hours of .129. See Table 4.18.

Given that for all workers, real wages decreased while hours worked increased, this also seems to suggest that the supply of workers increased as workers put value on the benefit of greater workplace safety. In the case of just private workers, the decrease in wages, but no change in hours, potentially indicates that while the supply of workers increased following state regulation of OSHA, there was a decrease in worker demand. This could occur if state regulation of OSHA increased employer's costs of hiring workers, particularly if regulation is better enforced or becomes more stringent. From the theory of mandated benefits, the fact that employment did not change significantly suggests that workers valued the benefit of greater workplace safety at its cost to employers.

For public sector workers, there was no change in real wages along with a significant decrease in hours worked. Given that there was no federal mandate of workplace safety prior to creation of state regulated OSHA programs, it is almost likely that employer costs increased afterwards, leading to a fall in worker demand in the public sector.

Table 4.18. Impact of State OSHA Laws on Hours Worked

	(1) All Workers	(2) Private Sector	(3) Public Sector
State Law/Certified	0.0615** (0.026)	0.00730 (0.035)	-0.129** (0.063)
Age	0.0189*** (0.00029)	0.0180*** (0.00033)	0.0249*** (0.00078)
Female	-2.554*** (0.0077)	-2.628*** (0.0091)	-2.013*** (0.020)
Black	0.0993*** (0.013)	-0.176*** (0.016)	0.669*** (0.031)
Hispanic	0.581*** (0.015)	0.440*** (0.017)	0.324*** (0.045)
Other - NonWhite	-0.228*** (0.019)	-0.248*** (0.022)	-0.293*** (0.047)
HS Degree	2.794*** (0.012)	2.787*** (0.013)	2.842*** (0.040)
Some College	2.433*** (0.012)	2.455*** (0.014)	2.420*** (0.040)
College Degree	3.968*** (0.014)	3.802*** (0.016)	4.486*** (0.041)
Advanced Degree	5.179*** (0.017)	4.840*** (0.022)	5.959*** (0.043)
Married	0.818*** (0.0078)	0.883*** (0.0090)	0.711*** (0.021)
Union Coverage	1.016*** (0.025)	0.699*** (0.039)	1.766*** (0.036)
Observations	4676322	3464730	654109
R^2	0.83	0.83	0.83

Note: The dependent variable in all regressions is number of usual weekly hours worked. Data are individual level observations, from the 1979-2009 CPS ORG. State, 2 digit industry and year fixed effects are included in all specifications. Coefficients that are significant at the .1, .05, .01 percent level are indicated with *, **, ***, respectively.

Table 4.19. breaks down the change in hours worked within the private sector. While hours worked has not changed significantly across most industry sectors, there have actually been increases in hours worked in the construction and wholesale trade industries. Given that construction is by far the most dangerous industry to work in, workers might have valued the increased safety from state certification of OSHA to an extent greater than cost to employers, leading to large increase in supply. Interestingly, there has been a significant fall in hours worked in the services sector after state certification of an OSHA program.

Table 4.19. Impact of State OSHA Laws on Hours Worked, by Private Sector

	(1) Agriculture	(2) Mining	(3) Construction	(4) Manufacturing	(5) Transp	(6) Whole. Trade	(7) Ret. Trade	(8) Finance	(9) Services
State Law/Certified	-0.409 (0.28)	0.353 (0.37)	0.361*** (0.11)	0.0200 (0.043)	0.0994 (0.15)	0.513*** (0.14)	-0.0765 (0.097)	-0.130 (0.12)	-0.277*** (0.077)
Age	0.00520** (0.0021)	-0.0438*** (0.0045)	-0.00195 (0.0012)	0.00432*** (0.00042)	0.00724*** (0.0018)	-0.0135*** (0.0014)	0.0663*** (0.00099)	-0.00938*** (0.0012)	0.00947*** (0.00064)
Female	-2.987*** (0.056)	-5.191*** (0.14)	-2.951*** (0.043)	-1.343*** (0.011)	-3.081*** (0.045)	-2.832*** (0.036)	-4.146*** (0.025)	-2.077*** (0.032)	-2.651*** (0.018)
Black	0.0363 (0.032)	-0.643*** (0.078)	-0.245*** (0.019)	-0.103*** (0.0069)	-0.443*** (0.031)	-0.201*** (0.025)	0.0782*** (0.018)	-0.189*** (0.023)	0.0606*** (0.011)
Education	0.670*** (0.027)	-0.160*** (0.049)	0.593*** (0.015)	0.474*** (0.0050)	0.176*** (0.021)	0.704*** (0.017)	1.466*** (0.013)	1.035*** (0.016)	1.140*** (0.0075)
Married	2.107*** (0.060)	0.397*** (0.12)	0.914*** (0.029)	0.532*** (0.011)	0.939*** (0.046)	0.884*** (0.037)	2.698*** (0.028)	0.290*** (0.032)	0.386*** (0.017)
Union Coverage	0.0500 (0.45)	-1.581*** (0.37)	0.866*** (0.13)	0.168*** (0.042)	-0.650*** (0.12)	0.683*** (0.20)	1.003*** (0.15)	0.949*** (0.19)	1.137*** (0.079)
Observations	130146	38472	204678	866297	167872	161028	584528	221431	1047455
R^2	0.78	0.82	0.90	0.94	0.83	0.89	0.72	0.86	0.78

Note: The dependent variable in all regressions is number of usual weekly hours worked. Data are individual level observations, from the 1979-2009 CPS ORG. State, 2 digit industry and year fixed effects are included in all specifications. Coefficients that are significant at the .1, .05, .01 percent level are indicated with *, **, ***, respectively.

4.10. Conclusion

The creation of state regulated OSHA programs provides an interesting study of the differential impacts of state versus federal enforcement. I find that state regulation of OSHA, as measured by certification of the state program, leads to an increased number of inspections per capita and citation of violations, as compared to federal regulation of OSHA.

Despite state regulated OSHA programs employing greater inspections and violations per capita, in the more recent period, state OSHA programs have not been associated with fewer nonfatalities from 1996-2008 or fewer fatalities from 2003-2008. This may suggest inefficiencies at the state level or regulatory capture/corruption of state inspections by local business interests.

I also explore the differences that exist between state OSHA programs, depending on the political climate in which they were created and the degree to which they differ from the federal program. I find that state programs generally experienced an increase in number of inspections per capita and violations per capita compared to federal programs, but do not have lower rates of nonfatal injuries or fatalities in the more recent period. It may be that current political climate is a better indicator of enforcement and more research will be needed on this topic.

Interestingly, state programs that adopted standards beyond those required by the federal statute are associated with a significantly lower number of fatalities, compared to states that only attain the minimum requirements of the federal law. This suggests that more stringent requirements are effective at improving workplace safety.

Evidence from the CA 2000 penalty changes suggests that increasing penalty maximums, as well as subjecting violations to criminal charges, may be an effective means of improving workplace safety. The California reform was associated with an increased number of inspections and violations, accompanied with a significant reduction in the number of fatalities following the penalty change.

Finally, I explore the impact of state regulation of OSHA on wages and employment. The increased use of inspections and issuance of violations following state regulation may have also increased costs for employers of maintaining safer workplaces, affecting the hiring cost of workers. There certainly appears to be a compensating differential for workplace safety, as wages fall significantly, particularly in the more dangerous industries, such as construction and manufacturing. However, the supply of workers also seems to have shifted out, as workers valued the benefit of greater workplace safety, leading to no change in employment, or hours worked, or even an increase in hours worked.

These findings can be particularly informative as OSHA reform continues to be a widely discussed topic. I hope to supplement my analysis with more detailed information on the types of violations issued and the magnitude of penalties, to assess whether there are differential issuance of violations and penalties by state regulated OSHA programs, compared to federally regulated programs. Further work remains to be done on the impact of voluntary compliance programs implemented by many states and the more recent reforms to increase the number of inspection officers.

BIBLIOGRAPHY

1st Session 98th Congress, *Senate Report No. 98-225* 1983.

Abrams, David S., Marianne Bertrand, and Sendhil Mullainathan, “Do Judges Vary in Their Treatment of Race?,” *Journal of Legal Studies*, (forthcoming 2012).

Alschuler, Albert W., “Sentencing Reform and Prosecutorial Power: A Critique of Recent Proposals for “Fixed” and “Presumptive” Sentencing,” *University of Pennsylvania Law Review*, 1978, 126, 550–577.

American Friends Service Committee, *Struggle for Justice: A Report on Crime and Punishment in America*, Hill & Wang, 1971.

Anderson, James M., Jeffrey R. Kling, and Kate Stith, “Measuring Interjudge Sentencing Disparity: Before and After the Federal Sentencing Guidelines,” *Journal of Law and Economics*, 1999, 42, 271–307.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson, “The Impact of Jury Race in Criminal Trials,” *Quarterly Journal of Economics*, 2012, 127, 1017–1055.

Austin, William and Thomas A. Williams III, “A Survey of Judges’ Responses to Simulated Legal Cases: Research Note on Sentencing Disparity,” *Journal of Criminal Law and Criminology*, June 1977, 69 (2), 306–310.

Autor, David H. and Susan N. Houseman, “Do Temporary-Help Jobs Improve Labor Market Outcomes for Low-Skilled Workers? Evidence from ‘Work First’,” *American Economic Journal: Applied Economics*, July 2010, 2 (3), 96–128.

Bartel, A.P. and L.G. Thomas, “Direct and Indirect Effects of Regulation: A New Look at OSHA’s Impact,” *Journal of Law and Economics*, 1985, 28 (1), 1–28.

Berman, Douglas A., “Foreword: Beyond Blakely and Booker: Pondering Modern Sentencing Process,” *Journal of Criminal Law and Criminology*, 2005, 95, 653–689.

Bradbury, John, “Regulatory Federalism and Workplace Safety: Evidence from OSHA Enforcement, 1981–1995,” *Journal of Regulatory Economics*, 2006, 29 (2), 211–224.

Fischman, Joshua B. and Max M. Schanzenbach, “Do Standards of Review Matter? The Case of Federal Criminal Sentencing,” *Journal of Legal Studies*, 2011, 40, 405–437.

—— and ———, “Racial Disparities under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums,” *Journal of Empirical Legal Studies*, (forthcoming).

Frankel, Marvin E., *Criminal Sentences: Law Without Order*, New York: Hill & Wang, 1973.

Freeborn, Beth A. and Monica E. Hartmann, “Judicial Discretion and Sentencing Behavior: Did the Feeney Amendment Rein in District Judges?,” *Journal of Empirical Legal Studies*, 2010, 7, 355–378.

Freed, Daniel J., “Federal Sentencing in the Wake of the Guidelines: Unacceptable Limits on the Discretion of Sentencers,” *Yale Law Journal*, 1992, 101, 1681–1754.

Gennaioli, Nicola and Andrei Shleifer, “Judicial Fact Discretion,” *Journal of Legal Studies*, January 2008, 37, 1–35.

Gertner, Nancy, “What Yogi Berra Teaches About Post-Booker Sentencing,” *Yale Law Journal Pocket Part*, 2006, 115, 137–141.

Gray, W.B. and C.A. Jones, “Are OSHA Health Inspections Effective? A Longitudinal Study in the Manufacturing Sector,” *The Review of Economics and Statistics*, 1991, 73, 504–508.

—— and ———, “Longitudinal Patterns of Compliance with OSHA Health and Safety Regulations in the Manufacturing Sector,” *Journal of Human Resources*, 1991, 26, 623–653.

—— and **J.M. Mendeloff**, “The Declining Effects of OSHA on Manufacturing Injuries: 1979 to 1998,” *NBER Working Paper 9119*, 2002.

- and **J.T. Scholz**, “Does Regulatory Enforcement Work? A Panel Analysis of OSHA Enforcement,” *Law & Society Review*, 1993, 27, 177–213.
- Hofer, Paul J., Kevin R. Blackwell, and R. Barry Ruback**, “The Effect of the Federal Sentencing Guidelines on Inter-Judge Sentencing Disparity,” *Journal of Criminal Law and Criminology*, 1999, 90, 239–321.
- Kuziemko, Ilyana**, “Should Prisoners be Released via Rules or Discretion?,” NBER Working Paper 13380, 2011.
- Leeth, John D. Kniesner Thomas J &**, “Abolishing OSHA,” *Regulation*, 1995, 46 (4).
- Maukestad, William J. and Charles Helm**, “Promoting Workplace Safety and Health in the Post-Regulatory Era: A Primer on Non-OSHA Legal Incentives that Influence Employer Decisions to Control Occupational Hazards,” *Kentucky Law Review*, 1989, 9.
- Morantz, Alison D.**, “Has Devolution Injured American Workers? State and Federal Enforcement of Construction Safety,” *Journal of Law, Economics & Organization*, 2009, 25 (1), 183–210.
- Mustard, David B.**, “Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts,” *Journal of Law and Economics*, 2001, 44, 258–314.
- Nagel, Ilene H. and Steven J. Schulhofer**, “A Tale of Three Cities: An Empirical Study of Charging and Bargaining Practices Under the Sentencing Guidelines,” *Southern California Law Review*, 1992, 66, 501–566.
- Posner, Richard A.**, “Judicial Behavior and Performance: An Economic Approach,” *Florida State University Law Review*, 2005, 32, 1259–1279.
- Rehavi, M. Marit and Sonja B. Starr**, “Racial Disparity in Federal Criminal Charging and its Sentencing Consequences,” January 2012. University of Michigan Law & Economics, Empirical Legal Studies Center Paper No. 12-002.
- Ruser, J.W. and R.S. Smith**, “The Effect of OSHA Records-Check Inspections on Reported Occupational Injuries in Manufacturing Establishments. Journal of Risk and Uncertainty,” *Journal of Risk and Uncertainty*, 1988, 1 (4), 415–435.
- Schanzenbach, Max M.**, “Racial and Sex Disparities in Prison Sentences: The Effect of District-Level Judicial Demographics,” *Journal of Legal Studies*, 2005, 34 (1), 57–92.
- and **Emerson H. Tiller**, “Strategic Judging Under the United States Sentencing Guidelines: Positive Political Theory and Evidence,” *Journal of Law, Economics, and Organization*, 2007, 23 (1), 24–56.
- and —, “Reviewing the Sentencing Guidelines: Judicial Politics, Empirical Evidence, and Reform,” *University of Chicago Law Review*, 2008, 75, 715–760.
- Scott, Ryan W.**, “Inter-Judge Sentencing Disparity After Booker: A First Look,” *Stanford Law Review*, December 2010, 63 (1), 1–66.
- Smith, R.S.**, “The Occupational Safety and Health Act: Its Goals and Its Achievements,” *American Enterprise Institute for Public Policy Research*, 1976.
- Stith, Kate**, “The Arc of the Pendulum: Judges, Prosecutors, and the Exercise of Discretion,” *Yale Law Journal*, 2008, 117, 1420–1497.
- and **Jose A. Cabranes**, *Fear of Judging: Sentencing Guidelines in the Federal Courts*, The University of Chicago Press, 1998.
- Tonry, Michael**, “Obsolescence and Immanence in Penal Theory and Policy,” *Columbia Law Review*, May 2005, 105 (4), 1233–1275.
- Ulmer, Jeffery T., Michael T. Light, and John H. Kramer**, “Racial Disparity in the Wake of the Booker/Fanfan Decision: An Alternative Analysis to the USSC’s 2010 Report,” *Criminology and Public Policy*, 2011, 10 (4), 1077–1118.

- United States Sentencing Commission**, *Measuring Recidivism: The Criminal History Computation of the Federal Sentencing Guidelines* May 2004.
- , *Final Report on the Impact of United States v. Booker on Federal Sentencing* 2006.
- , *Demographic Differences in Federal Sentencing Practices: An Update of the Booker Report's Multivariate Regression Analysis* 2010.
- , *Report to Congress: Mandatory Minimum Penalties in the Federal Criminal Justice System* October 2011.
- Viscusi, W.K.**, “The Impact of Occupational Safety and Health Regulation,” *Bell Journal of Economics*, 1979, 10 (1), 117–140.
- , “The Impact of Occupational Safety and Health Regulation, 1973-1983,” *RAND Journal of Economics*, 1986, 17 (4), 567–580.
- Weil, David**, “IF OSHA is So Bad, Why is Compliance So Good?,” *RAND Journal of Economics*, 1996, 27, 618–640.
- , “Assessing OSHA Performance: New Evidence from the Construction Industry,” *Journal of Policy Analysis and Management*, 2001, 20 (4), 651–674.
- Welch, Susan, Michael Combs, and John Gruhl**, “Do Black Judges Make a Difference?,” *American Journal of Political Science*, 1988, 32, 126–136.
- Wheeler, Stanton, Kenneth Mann, and Austin Sarat**, *Sitting in Judgment: The Sentencing of White-Collar Criminals*, Yale University Press, 1988.